

## Introduction

Evolutionary theory has, seemingly since its inception, called for careful philosophical interpretation. Some reasons for this are straightforward enough: It brings into question our own place in the universe (Hull, 1986; Machery, 2008) and threatens the cogency of traditional theories of ethics (Street, 2006) and epistemology (Bradie, 1986). Other concerns fall more squarely within the province of the philosophy of science. As soon as Darwin had set the *Origin* to paper, it was quickly recognized that this was, at the very least, an unusual scientific theory. Its argument rests on a deft combination of analogy from sophisticated but still “outsider” knowledge like that of agricultural breeders or pigeon fanciers (Sterrett, 2002), cutting-edge geological theory supported by Darwin’s own observations on the Beagle voyage (Hodge, 1983), and connections with a surprisingly rich philosophical background (Sloan, 2009). This did not look like ordinary nineteenth-century science, as Darwin was quickly told by figures as diverse as the physicist-astronomer John Herschel (Pence, 2018), the engineer Fleeming Jenkin (1867), and the geologist Adam Sedgwick (1860).

What kinds of worries might one raise when it comes to the nature of evolution qua scientific theory? First, evolutionary theory’s relationship to the evidence that supports it – even with access to contemporary evolutionary modeling, theories of probability, and statistical inference – is quite convoluted (see, e.g., the debate over probability and statistics in Sober, 2008). Darwin relied upon both contemporary observations that could not necessarily be extrapolated cleanly to the case of long-term population change, and on the fossil record, which is often taken to ground very different kinds of explanatory enterprises (Currie, 2019). This basis has been supplemented by a panoply of further sources in the intervening century-and-a-half, ranging from the biochemical fundamentals of DNA to large-scale ecological observations that Darwin would have thought impossible. Such contributions increase confidence in our conclusions about changes in life on earth but certainly do nothing to make the theory–evidence relationship easier to comprehend.

The structure of Darwin’s presentation of natural selection was also quite novel.<sup>1</sup> Building a conception of the generation of adaptations predicated on what we would now call a process of biased sampling across generations (Hodge, 1987), with only the rudiments of statistical theory having yet been developed, and very little application even of those to problems in the biological

---

<sup>1</sup> The *Origin* also includes Darwin’s argument for common ancestry, which is less innovative in this sense, and was also much more rapidly accepted by biologists of his day.

world, is quite a feat indeed (Sheynin, 1980; Hacking, 1990). It left a number of open questions about how we should conceive of the organisms, populations, and species that Darwin's theory described. The modern answers to those questions would only start to take shape after nearly a century of debate and discussion concerning whether and how to integrate probability and statistics into our understanding of evolutionary biology (Pence, in press).

One of these open questions concerned the relationship between the process of heredity or inheritance and the kinds of explanations produced by natural selection (Gayon, 1998). How much knowledge of patterns of variation and its transmission from parents to offspring was really necessary in order to provide a suitable grounding for our conclusions about evolutionary change? Darwin had only a fairly weak theory to offer, and while Mendel's work appeared to provide an alluring possibility to complete the picture, figuring out exactly how Mendelian results – data, for one thing, only about the ratios of offspring produced by a single type of mating event, not natural populations – were supposed to provide explanations of Darwinian, populational, selective phenomena was by no means immediately clear.

A second such question – or perhaps better, a refinement of the first – addresses the importance of the underlying material basis of heredity for evolutionary theory. By the early years of the twentieth century it was increasingly accepted that, somehow or other, chromosomes were intimately involved in the transmission of characters. But what kind of account of them, if any, would be needed to finish the evolutionary story? Were Mendel-style phenomenological patterns in the production of offspring enough, or did we need to tie those down to their underlying physico-chemical details? Or, to reframe things from today's perspective, how much do the nature and transmission of DNA really matter to evolutionary biology? Unlike the first question, which seems somewhat anachronistic from our perspective, this cluster of issues is far from resolved, with debate still raging about the nature of mutations and their importance for evolutionary change (Merlin, 2016; Stoltzfus and Yampolsky, 2009; Stoltzfus, 2019).

Third, we have the related fact that variation seems to be a distributional phenomenon. Except in relatively isolated cases like the apparently clear-cut traits of Mendel's peas, variations are often normally distributed (Galton, 1889), and detecting natural selection in the wild was soon taken to involve (at least, by those who believed that there was such a thing as natural selection to be detected at all) examining the changes in these distributions over time (Weldon, 1893). This presented yet another disconnect between theorizing about natural selection and the "traditional" life sciences, whose morphological approaches seemed to have had no need for such methods thus far.

Fourth and finally – and intimately related to the last open interpretive question – we have the issue that most directly concerns my work here, one that, again, has remained with us to the present day. Natural selection is, in the end, a description of a *probable* outcome (namely, adaptation to environmental conditions). It offers us no guarantees: Fitter organisms are only, all else equal, more likely to leave more offspring than the less fit, and fitter traits are more likely to spread than the less fit. In that sense, it is unlike either a universal law of nature, taken to necessarily imply that, say, for any two bodies whatsoever in the universe, the force of gravitational attraction between them is and must be  $Gm_1m_2/r^2$ , or an inductive generalization inferred from observing a vast number of, say, white swans, ever on the lookout for a rogue disconfirming instance.

If this is so, then what exactly *is* natural selection? How should it be understood? As Hodge has described it, biologists seeking to define natural selection in textbooks on evolution readily agree upon a partial set of necessary conditions for natural selection to act: “variation, heritability of variation, and differential reproduction of heritable variation.” All also recognize “that some further condition is necessary, so that once this is given there will be a set of necessary conditions that are jointly sufficient” (Hodge, 1987, p. 250). Put differently, we need variation, inheritance, and differential reproduction, but we must also delineate the right sort of differential reproduction – the selective kind – from others due to other kinds of influences on the evolutionary process – like that due to drift, migration, and so forth.

It is in making this final distinction that we reach a point of divergence among the textbooks. Some authors turn to statistical qualifications, namely that the differential reproduction at work in natural selection must be

“consistent” or “systematic” or “nonrandom,” all terms with no peculiarly biological content and drawn often from the terminology of statistics, while other authors insist that the differential reproduction must be due to differences in “fitness” or “adaptation,” terms characteristic of the biologists’, even formerly the theologians’, lexicon, terms with an apparent teleological import. (Hodge, 1987, pp. 250–251)

That is, some authors appeal only to the statistical nature of the patterns of the variation on offer. Others argue that natural selection is picked out by its effects; it is the sort of process that produces increases in fitness or produces adaptations (it is this sort of explanation that Hodge rightly notes has a teleological flavor).

These two common moves are to be contrasted, he argues, with a third choice – namely, “an explicit definitional insistence on causation itself,” for “differential reproduction in selection is distinguished from any in

[non-selective cases like] drift by its causation” (Hodge, 1987, p. 251), in particular, by whether or not the variation at issue is causally relevant to the differential reproduction being explained. Here we focus neither on the statistical nature of differential reproduction nor on the outcomes that result, but on the (causal) nature of the *process* that connects variation to differential selection.

In short, Hodge sees three distinct ways in which we might interpret natural selection. We could attempt a purely statistical characterization, trying to specify in the mathematical properties of sampling from populations those that pick out natural selection, as opposed to other evolutionary sampling processes. Alternatively, we could look retroactively for the signal of natural selection in terms of its effects – though this runs a serious risk of reintroducing a sort of teleological explanation that Darwin has often been taken to have rendered unnecessary in evolutionary thought. Lastly, we could explore the nature of the causal connections underlying natural selection itself. If selection is constituted by a subset of the causal influences that impinge upon organisms in the wild, perhaps that subset shares a common feature that would let us pick out selective from nonselective change.

This Element is, essentially, an examination of the last several decades of efforts to describe the underlying causal structure of the theory of evolution by natural selection. As we will see, since the problem was first given a clear formulation in the mid-1980s, there has ensued constant and fairly heated debate concerning the appropriate way to understand the causal structure of selection. In Section 1, I will introduce what is now often referred to as the “causalist/statisticalist debate,” with a particular focus on attempting to disentangle the question of the *causal structure* of natural selection from a whole host of other concerns that have since become embroiled in it. Section 2 introduces a novel way of presenting the debate, via diagramming these causal structures. The lens for introducing this structure will be the statisticalist view, but a more general version of that structure can readily allow us to represent the underlying causal commitments of several varieties of causalism, which is the task of Section 3. Finally, in Section 4, I’ll use this new presentation of the players in this arena to try to make some advances, connecting to work on causal exclusion arguments in the metaphysics of science and philosophy of mind and on universality in the philosophy of physics.

While I certainly have my own position in this debate, having published on the causalist approach to selection, the aim of this Element is not to offer decisive arguments in favor of one position or the other. Rather, I hope here to be responding to a general sense of stagnation surrounding the discussion – arguments are drawn up, responses written, and yet the positions of few authors

seem to engage in anything like systematic response to one another. The hypothesis underlying my approach is that this is in part, at least, due to our having pervasively talked past one another. One way, then, to try to repair the failure of our work to connect is to take up the task of clarifying exactly what is at stake. As I will argue in the next section, there are a plethora of interesting philosophical questions that have been lumped together under the banner of the causalist/statisticalist debate, and many responses to these, in turn, have been forced into the boxes of one “camp” or the other. My goal here is to take a first step toward separating such questions, to pick up one that I believe has been broadly under-studied in the literature, and to show that understanding it can provide profitable opportunities for connection with other areas in the philosophy of science.

One more caveat is due before I continue. I am explicitly narrowing my scope in what follows to a presentation of *natural selection*, with a few unavoidable detours into concepts of fitness, populations, and traits. Of course, to do this is to neglect philosophers’ other favorite evolutionary process, genetic drift. It also sets aside the other half of the “big four” factors that would frame an introductory course in evolutionary biology, mutation and migration. It omits as well the wider array of influences on evolutionary trajectories whose importance we are increasingly coming to appreciate as calls for an “extended” evolutionary synthesis become more widespread (Pigliucci, 2007), including niche construction, development, and genotype-environment interaction (see, e.g., Uller and Laland, 2019). I am fully aware that some readers may be inclined to throw such an “old-fashioned” approach as the one I’m pursuing here across the room (if reading a paper copy). That said, I think the choice to restrict my attention to natural selection is defensible for two reasons. First, the kind of structural features of selection that I’ll point out here will be equally relevant for any evolutionary process – that is, anything giving rise to population change as a result of aggregated, individual-level events.<sup>2</sup> Selection will thus, at worst, prove useful as an analogy for these other cases. Second, as we will see, the causal structure of selection has been the site of particularly heated debate in recent years, to which we should now turn.

## 1 The Contemporary Debate over Causation in Natural Selection

It is no simple matter to understand causation in evolving systems. Individual organisms have, since Aristotle, been paradigmatic examples of causal agents,

---

<sup>2</sup> When discussing various positions in the causalist/statisticalist debate in the first few sections, I will use “organism” and “individual” interchangeably, as has been common practice in this literature; I will return to this question and clarify the notion of “individual” important for my own analysis in Section 2.3.

but the onward march of the life sciences has both broken these organisms down into their component parts – organs, tissues, cells, proteins, and nucleic acids – and combined them into successively larger groupings – trait groups, populations, species, and ecosystems. Insofar as these are all descriptions of the same underlying organic phenomena, philosophers of biology have an interesting interpretive task in front of us: How do we harmonize these descriptions, and which ones ought we to prefer in which kinds of circumstances? Where is the real “causal action” in evolutionary theory? From Ernst Mayr’s “population thinking” (1961, 1976), to the periodic heralding of the “return” of entities like groups (Sterelny, 1996) or organisms (Nicholson, 2014), to the focus of evolutionary theory, this issue has remained important to a variety of philosophers interested in the interpretation of natural selection.

### 1.1 The Received View

As classically crystallized by Elliott Sober’s *The Nature of Selection* (1984), a “received view” of the fundamental causal structure of evolutionary theory was developed in the first few decades of contemporary philosophy of biology. On this picture, natural selection is a causal process driving evolutionary change. The strength and direction of this process are, in turn, described by Darwinian fitness. The proper way to understand this fitness property spawned its own literature, internal to the received view (Mills and Beatty, 1979; Beatty and Finsen, 1989; Brandon, 1990; Sober, 2001), too complex for me to engage with in great detail here. In short, the idea is that fitness describes not some simple, phenotypic or demographic property of an organism (like its speed, strength, or the number of offspring it has had to date), but rather its propensity to survive and reproduce – that is, how likely it is to have various numbers of offspring. Just as a fragile glass is likely to break into more pieces when dropped than a sturdy one, a fit organism is likely to have more offspring than an unfit one, holding all else equal. Fitness, on this approach, summarizes a vast number of the causally relevant properties of an organism that, when summed together in the right way, describe the probability of various future possible outcomes of natural selection.<sup>3</sup>

Natural selection is then taken to be a causal process operating on populations, the direction and strength of which are governed in part by fitness,

<sup>3</sup> Notably, this describes only the earliest forms of the received view; as we will see below, a number of more sophisticated versions of causalism are now on offer.

which is the propensity of an organism to survive and reproduce.<sup>4</sup> Darwin's key insight, on this reading, was to see that if an organism is more likely to survive and reproduce (or if a character trait makes the organisms that bear it more likely to survive and reproduce) then, in the long run and all else equal, that organism (or trait) will probably increase its representation in the population. Contra an objection most commonly ascribed to Karl Popper (1974), such a theory is not tautologous, because there is no guarantee of evolutionary success to be found here – the fittest are only *more likely* to survive.<sup>5</sup>

This view has a few convenient advantages. As Sober argued in his initial presentation (though the story is now recognized to be significantly more complicated than this; see Sober, 2013), it can cleanly separate the notions of “selection of” and “selection for.” Some traits are directly promoted by natural selection, while others increase in prevalence only because they are correlated with traits that improve fitness. To borrow a classic example, hearts are evolutionarily favored because they pump blood, not because they make thumping sounds, even if the properties of pumping blood and making thumping sounds are in fact coextensive. There has thus been *selection for* pumping blood, while there has only been *selection of* making thumping sounds. The received view gives us the tools we need to distinguish these two cases. If a trait is associated with an increase in fitness in a counterfactually and causally robust kind of way, which we can discover by examining the propensity that describes fitness, then it is being selected for; if we can see that the trait is not in fact causally relevant to an organism's success, then it is merely benefiting from “selection of.” We could imagine noiseless hearts that still contributed to survival and reproduction; we cannot imagine pumpless thumpers that do the same.

We can also use this approach to identify the differing causal impacts of the various “forces” or “factors” that contribute to evolutionary theory. Since selective evolutionary change is clearly defined, then other sorts of change – especially that due to genetic drift, which has been an important topic for philosophers of biology for other reasons (Hodge, 1987; Millstein, 2002, 2008) – can be examined comparatively. Net evolutionary change is thus considered to be the result of the cumulative impacts of each of these features, like selection, migration, mutation, drift, and so forth. This kind of

---

<sup>4</sup> This slippage between the organismic and population levels would go on to be one of the key drivers behind the development of more sophisticated causalist views.

<sup>5</sup> It is worth emphasizing that Popper himself eventually abandoned this objection (Elgin and Sober, 2017).

compositional approach (whatever ontology might in fact underlie it) matches extremely well with the way in which biologists talk about the process of evolution (Luque, 2016; Pence, 2017).

## 1.2 The Statisticalist View

The philosophical situation changed radically, however, in 2002, with the publication of two articles directly challenging this received view (Matthen and Ariew, 2002; Walsh et al., 2002). These articles inaugurated what has since become known as the *statisticalist* interpretation of evolutionary theory, by contrast with the received view, now known as the *causalist* position. The statisticalist approach is complex, and I will present it more thoroughly later. But it can be summarized roughly as follows: First, the statisticalists argue that natural selection and genetic drift (other factors in evolution are rarely considered) are merely convenient summaries of the genuinely causal events taking place in the lives of individual organisms. When organisms eat, fight, mate, and die, these causal events are what powers evolution. We, as theorists, then make the decision to abstract away from these fine-grained details and build evolutionary explanations using terms like selection, fitness, and drift.

This abstraction has several important effects, the statisticalists argue. First, the assignment of selection or drift to a given population change is relative to particular observers and their interests, and thus there is no single fact of the matter about which of these are occurring “out there” in the world (Walsh et al., 2002, p. 467) – which entails that natural selection and genetic drift (in the sense of the received view, at least) do not exist “out there” in the world, either. In this sense, natural selection is a bit like the Dow Jones Industrial Average: potentially extremely useful for us in understanding the stock market, but not an independent feature of the world to be discovered, and intimately connected to our subjective interests in continuing to measure it.

Second, statisticalists argue that these abstractions cannot be causal. At a sufficient level of generality (after, that is, enough abstraction), we are left only with a priori statistical identities. The effects of natural selection or genetic drift, that is, are only mathematical consequences of understanding populations as certain kinds of statistical objects. Multigenerational sampling of various sorts, as is required simply in writing down a formalism to represent an evolving system, entails certain kinds of changes over time, which directly correspond to the mathematical formalisms of selection and drift. Because these entailments are analytic, they are not causal.

Third, and as an immediate extension of this last point, models of natural selection are *substrate-neutral* – that is, they make no reference whatsoever



to what it is that is being selected or what it is that is drifting.<sup>6</sup> The very arrangement of a population (that is, into individuals bearing traits with inheritance and differential success) will imply the existence of some kinds of change (namely, an analogue to selection, where better-reproducing types will be better represented in the future, and an analogue to drift, where the most probable result will be disrupted due to sampling error thanks to finite population size). If we arranged anything, from marbles to coins to elk, in those kinds of structures over time, we would see processes that looked very much like selection and drift in the “populations” that resulted.<sup>7</sup>

The extent to which this stands as a challenge to the received view should be relatively clear: Natural selection and genetic drift are not causal processes; fitness (at least, as we will see below, the kind of fitness actually relevant for evolutionary change) is not a causal property, much less a propensity; evolutionary explanations are subjective abstractions, not objective explanations of population phenomena; natural selection is indeed an a priori claim about certain kinds of statistical assemblages. Darwin should be viewed, they argue, as having described facts about the various causal events that individual organisms are involved in (instances of success or failure in the struggle for life) and their long-term consequences for individual lineages, which are *not* the target of the mathematical models of contemporary population genetics.

Such a challenge was rapidly picked up by the causalists, and what followed has been nearly twenty years (and counting) of ever-ramifying debate, encompassing more and more aspects of evolutionary theory. We have seen contention over the very definitions of natural selection and genetic drift (Millstein, 2002; Pfeifer, 2005; Brandon, 2006; Plutynski, 2007; McShea and Brandon, 2010; Ramsey, 2013b; Strevens, 2016), with a particular emphasis on a distinction between accounts of those that treat them as processes working in the world, and other accounts that treat them as identifying particular outcomes or results within populations (Brandon, 2005; Millstein, 2005; Millstein et al., 2009).

This has coincided with renewed debate over the nature of fitness. Authors have disputed whether or not fitness should be taken to be a property of

<sup>6</sup> This claim is not exclusively made by the statisticalists – for instance, it is very important to Sober (1984), a paradigmatically causalist work. It nonetheless forms an important part of the statisticalist argument, and so I will evaluate it as such in what follows.

<sup>7</sup> The presentation here – as with all my presentations of the statisticalist position in what follows – largely draws upon the recent “programmatic” piece by Walsh et al. (2017). I will, for reasons of space, be forced to pass over some important instances of differences of opinion under the broader statisticalist umbrella.

organisms, traits, or populations (Abrams, 2012a; Sober, 2013; Pence and Ramsey, 2015), whether the propensity interpretation itself is subject to a variety of proposed counterexamples and what changes to it would be necessary to resolve them (Ramsey, 2006; Abrams, 2009b; Otsuka et al., 2011; Pence and Ramsey, 2013; Ramsey, 2013c), and, perhaps most fundamentally, whether a causal property of fitness as proposed by the propensity interpretation can possibly fill the role that is demanded of it in the first place (Ariew and Lewontin, 2004; Bouchard and Rosenberg, 2004; Abrams, 2009a; Ariew and Ernst, 2009; Walsh, 2010; Ramsey, 2013a; Millstein, 2016; Triviño and Nuño de la Rosa, 2016).

Arguments over the nature of selection, in turn, have invoked the literature on causal processes, as well as Newtonian forces, in an attempt to clarify the locus of causation (Stephens, 2004; Millstein, 2013; Hitchcock and Velasco, 2014; Earnshaw, 2015; Luque, 2016; Pence, 2017). Note that this isn't quite the same as the classic "units of selection" debate (Okasha, 2006) – for one might believe that it is individual organisms that are being selected (that are the members of "Darwinian populations" sensu Godfrey-Smith, 2009), while still arguing that selection as a process acts upon populations. The question is more closely related to general considerations of supervenience and the relationship between organism-level, trait-level, and population-level causal facts (Reisman and Forber, 2005; Shapiro and Sober, 2007). As we consider how these questions are expressed in models of population genetics, we further have to understand the role of such mathematical models in the inferences drawn by population genetics, a topic that has been carefully explored by Jun Otsuka (2016, 2019).

Finally, one strand of the debate has looked at explanation as an approach to understanding the character of selection (putting a more epistemic gloss on a debate that has usually been framed in metaphysical terms), usually either pointing out the extent to which evolutionary explanations are sensitive to the context in which they are produced (Walsh, 2007), or in an effort to construct a new, noncausal sort of explanation that natural selection might involve (Matthen, 2009; Matthen and Ariew, 2009; Ariew et al., 2015).

There are, then, at least five different clusters of questions that all travel together under the banner of the "causalist/statisticalist debate." Placed roughly in order of decreasing proximity to biological practice, we have the following:

1. How should we define natural selection and genetic drift? Are they to be considered as processes acting upon populations, or population-level outcomes, or statistical identities? If they are processes, are they causal processes? Newtonian forces? Something else?