Forthcoming in *The British Journal for the Philosophy of Science*
https://doi.org/10.1086/728261

Relative Significance Controversies in Evolutionary Biology

Katie Deaven

Abstract

Several prominent debates in biology, such as those surrounding adaptationism, group selection, and punctuated equilibrium, have focused on disagreements about the relative importance of a cause in producing a phenomenon of interest. Some philosophers, such as John Beatty have expressed scepticism about the scientific value of engaging in these controversies, and Karen Kovaka has suggested that their value might be limited. In this paper, I challenge that scepticism by giving a novel analysis of relative significance controversies, showing that there are three forms they can take. I argue that these controversies can have significant epistemic upshots, in that they help scientists form predictions about new instances of the phenomenon of interest. Finally, using two historical examples, I show how engaging in these controversies can improve our understanding of causal relationships.

1. Introduction

Several prominent debates in biology, such as those surrounding adaptationism, group selection, and punctuated equilibrium, have focused on disagreements about the relative importance of various causes in producing a phenomenon of interest. Yet, many of the events for which a causal explanation is appropriate are the result of a dizzying number of causal factors. Furthermore, there are a variety of ways in which one cause might be more important than another. Why are some causes considered more important than others? This is the relative importance question (RIQ); those who have responded to the RIQ often take one of two approaches.

One approach simply involves spot-lighting certain causes. Some causes are selected and cited as 'the' causes of a phenomenon of interest, while others are backgrounded and treated as 'enabling conditions'. For example, in explaining the lighting of a match, it is the striking of the match that is often treated as *the* cause despite there being a second, equally necessary cause of the match lighting—oxygen being present. Some philosophers have questioned whether there is an objective rationale behind such selections or if they are merely the result of arbitrary, pragmatic, and/or non-scientific considerations (Mill [1874]; Lewis [1986]). This challenge has received significant attention in recent years, with several philosophers arguing that the spot-lighting approach to answering questions about relative importance is driven by objective considerations and improves our understanding of causal relationships (Waters [2007]; Woodward [2010]; Franklin-Hall [2015]; Ross [forthcoming]).

A second approach to evaluating the relative importance of causes involves quantifying them. Here, the question of interest is, 'Is C_1 or C_2 more significant for E's occurring?' rather than 'Which of C_1 or C_2 should be a cited cause of E?' This second approach generates relative significance controversies, which are common in evolutionary biology. Consider the dispute among biologists concerning speciation. Speciation events are classified as *allopatric* when they are the result of a geographic barrier that divides a single population into two subpopulations. This division is followed by different mutations arising within each subpopulation, after which drift and selection operate on different mutations to eventually produce reproductive isolation, from which the biological species concept concludes that there are now two species. A second kind of speciation process—sympatric speciation—occurs when a new species arises without a geographic barrier. The evolution of the cichlids of Lake Malawi is a classic example of this sort of speciation: strong selection pressures resulted in the rapid adaptive radiation of a single parent species into multiple sexually isolated lineages without geographic isolation (McCune and Lovejoy [1998]). Biologists have argued about the relative importance of each model for decades. When sympatric speciation was first proposed, doubt was cast on the idea that this kind of speciation could occur at all (Maynard Smith [1966]). Since then, sympatric speciation has been shown not only to be a theoretical possibility but to account for many speciation events that were once thought to be allopatric (Tregenza and Butlin [1999]).

This example illustrates one way in which causes can be quantified: the controversy is about the *frequencies* of the different causes, which yields claims about the relative importance of causes (e.g., geographic barriers are the most important cause of speciation). Scientists engaged in the speciation controversy are interested in answering this question: Of all speciation events, how often were they brought about by geographic separation (as opposed to some other factor)? This second approach to thinking about causal importance has also been criticized. Beatty ([1997]) questions whether the frequencies of causes in evolutionary biology yield anything ‘interesting’ or ‘telling’. He suggests that while these controversies are pursued for socio-political, financial, and pragmatic reasons, frequency data are contingent and evolving and will, therefore, not yield the kind of claims about causal importance that are worth knowing about (p. S440-441).[1] In response, Kovaka ([2021]) argues that these frequency controversies are indirectly valuable even if the questions at the heart of the controversy are not themselves scientifically interesting; scientists can use frequency data gained by settling the controversy to make inferences about the scope of causal explanations, which can provide explanatory insights.

In what follows, I respond to the scepticism about the quantifying approach to RIQ and argue that relative frequency claims can help scientists form predictions about new instances of the phenomenon of interest and improve our understanding of causal systems in evolutionary biology.[2] This is because relative frequency controversies stand in an important relationship to another kind of relative significance controversy. A second interpretation of the relative significance question—Is C_1 or C_2 more significant for E's occurring?—asks what the stronger of the two causes is. Causal efficacy, I contend, concerns the extent to which causes change the probability of their effects. While many relative significance controversies look like they are just investigations of relative frequencies, they often involve efficacy claims, too. By elucidating this relationship between relative frequency controversies and relative efficacy controversies, I show that frequency data are also valuable in that they sometimes yield information about the efficacy of causes and drive controversies of this kind forward. My arguments thus show that the quantification approach to RIQ can be defended against the kind of scepticism exemplified in Beatty ([1997]) in a way that is more robust than the qualified defence presented by Kovaka ([2021]).

[1] See Dietrich ([2020]) for an extended analysis of the interactions between social values and scientific reasoning that take place in relative significance controversies.

[2] Observe that these two approaches to RIQ are distinct: once the question of 'genuine' versus background causes is settled, there could still be a relative significance question about the causes that have been appropriately selected as the 'genuine' causes. Similarly, questions about the significance of causes—their frequency or their efficacy—could be asked about both the background and 'genuine' causes of some phenomenon. See Franklin-Hall ([2015]) for a second way of distinguishing frequency questions from questions about the selection of causes.

The organisation of this paper is as follows. I start by presenting a taxonomy of different kinds of relative significance controversies and analysing the components and the sets of conditions that give rise to each of them (Section 2). Then, I explicate the sceptical challenges raised by Beatty, Kovaka's response to them, and defend one epistemic upshot of engaging in relative frequency controversies (Section 3). In Section 4, I defend my claim that the relationship between frequency and efficacy claims has improved our understanding of causal systems—the second epistemic upshot—using two historical examples in evolutionary biology.

2. Analysing Relative Significance Controversies

Relative significance controversies arise when an explanandum is explained by more than one cause. This can happen in two ways: a single type of explanandum is sometimes caused exclusively by one cause and sometimes exclusively by another, or a single explanandum is the result of a multicausal process. In either case, the relative contribution of each cause lies at the heart of the ensuing controversy. These controversies are not only a matter of interest for philosophers; scientists engage in these controversies when they investigate the 'importance' of a cause in producing a phenomenon of interest.

Notice that the debate concerning speciation appears to be an example of the first kind of controversy in which there is only a single cause of every instance of the explanandum; each speciation event is either allopatric or sympatric.[3] Research on the evolution of dispersal patterns among terrestrial vertebrates illustrates what the second kind of controversy looks like. This research typically focuses on three distinct, but potentially concurrent, influences: competition for mates, competition for environmental resources, and inbreeding avoidance. Dobson and Jones ([1985]) argue that more than one of these factors influences a given dispersal pattern event. Thus,

[3] The reality might be more complicated than this; I return to this question in Section 2.2.

investigating the dispersal pattern of a species involves assessing the strength of each of these causes. An experiment on the causal efficacy of these factors in the evolution of dispersal patterns might involve manipulating the amount of food available in the natal area, the mating opportunities available to the individuals that leave the natal area, and the level of relatedness between individuals. Researchers would then determine the respective strength of each cause by analysing the changes in the observed dispersal pattern given each of these manipulations.

Observe that here, scientists are not concerned with the frequency of outcomes brought about by each of these causes. Rather, asking about the relative significance of the causes of dispersal amounts to asking: Which of these causes—competition for mates, competition for resources, or inbreeding avoidance—is the biggest difference-maker to the dispersal pattern? In this case, the cause's 'importance' consists in how efficacious it is compared to other causes. To distinguish these two ways of quantifying causes, I will use 'relative frequency controversy' to pick out controversies concerning the frequency of outcomes brought about by one or another of a set of singleton causes and reserve 'relative efficacy controversy' to pick out controversies about the efficacy of different causes that simultaneously impinge on a given effect. In what follows, I provide a taxonomy outlining the conditions under which these kinds of questions may be asked and consider examples of each.

2.1 Taxonomy of relative significance controversies

Given the distinction between causal frequency and causal efficacy, there are four exclusive and exhaustive situations to be explored. One of these may be eliminated easily. When a single instance of the explanandum is the result of a single cause, the question concerning the relative importance of the cause does not arise. The remaining categories are represented in Table 1 below.

Table 1: Taxonomy of Controversies

	One Cause of Each Instance of the Explanandum	Two or More Causes of Each Instance of the Explanandum
One Instance of the Explanandum		No Relative Frequency Controversy, Relative Efficacy Controversy
Two or More Instances of the Explanandum	Relative Frequency Controversy, No Relative Efficacy Controversy	Relative Frequency Controversy, Relative Efficacy Controversy[4]

The second example above concerning the evolution of dispersal patterns illustrates the upper-right cell in Table 1; it is a case where there are two or more causes that explain a single instance of the explanandum—namely, the evolutionary history of the observed dispersal pattern in a single population. Here, what it means to investigate the importance of a cause is to learn which cause contributed more to the occurrence of the explanandum—that is, which cause is more efficacious. One way of understanding the notion of causal efficacy is by thinking about the difference that causes make to their effects.

To see what it means for causes to differ in how much they contribute to the occurrence of an effect, consider the following experiment that researchers investigating the evolution of dispersal patterns might run. Suppose researchers are investigating the dispersal pattern of a population of black-tailed prairie dogs. They observe that adult male prairie dogs that have successfully reproduced disperse from the colony when their daughters mature, and this behavioural strategy (leaving the natal area) increases as the number of closely related individuals in the natal area increases. This suggests that the males are dispersing because doing so reduces inbreeding. However, suppose they also suspect this dispersal pattern is influenced by available food sources. To test the respective strengths of these influences, researchers will inquire about

[4] There is a second variation of this fourth box where there are two or more explanandum events and for some of these events, there is one cause for each explanandum, and for others, there are two or more causes for each explanandum.

the conditional probability of the observed dispersal pattern given conditions that would amplify or diminish the behavioural strategy.

Table 2: Conditional Probabilities of Observed Dispersal Pattern Given Manipulations of the Causal Influences

Pr(Dispersal Pattern Observed | —)		
	Individuals are closely related	Individuals are not closely related
Abundance of food sources	w	x
Insufficient food sources	y	z

In Table 2 above, each of the variables in the four cells has a value between 0 and 1 and represents the probability of the observed dispersal pattern given the respective conditions outlined. Assuming that both the level of food sources and the number of closely related prairie dogs in the natal area are efficacious, researchers will obtain data that supports horizontal ($w>x$ and $y>z$) and vertical ($w>y$ and $x>z$) inequalities between pairs of probabilities.[5] The quantity $w-x$ represents the causal contribution of relatedness, while the quantity $w-y$ represents the causal contribution of food availability. Suppose researchers discover from these experiments that $w-x>w-y$; one can infer from these values that the prairie dog population is more sensitive to changes in the number of closely related individuals than changes in food sources. From this, researchers might hypothesize about the selective forces that led to the evolution of that behavioural strategy. The results of this test would then support the hypothesis that inbreeding avoidance is a stronger influence on the observed behaviour than competition for resources. In other words, inbreeding

[5] It may be the case that researchers discover that the two causal factors are non-additive, and there is an interaction term in the top left cell. Since I am only comparing the values of y and x (the bottom-left and top-right cells), my argument does not depend on additivity.

avoidance is more efficacious than competition for food sources with respect to this dispersal pattern.[6]

This is not a novel approach to assessing causal efficacy.[7] Sober ([1988])—inspired by Lewontin's ([1974]) discussion of norms of reaction as a tool for disentangling how genes and environment affect phenotypes in a population—argues that causal responsibility for a person's height, given both her genes (G_1) and the environment she lives in (E_1), can be assessed using counterfactuals that address: (i) how tall she would be if she had different genes (G_2), but lived in the same environment (E_1), and (ii) how tall she would be if she lived in a different environment (E_2), but had the same genes (G_1) (p. 309). According to Sober ([1988]), to assess causal efficacy, the person's actual height given by $<E_1, G_1>$ should then be compared to the person's height based on the two counterfactuals. Suppose there is a larger difference between the person's actual height and her height in the counterfactual where her environment is different but her genes are held fixed $<E_2, G_1>$ than there is when comparing the person's actual height to the counterfactual situation where her genes are changed and her environment is held fixed $<E_1, G_2>$. Then, this would mean that the environment is the more powerful influence. Like the case of dispersal patterns, Sober is interested in comparing the conditional probability that the person's height obtains given a manipulation to one of the purported causes with the conditional probability that the person's height obtains given a manipulation to the other cause.[8]

---

[6] In order for researchers investigating the evolution of dispersal patterns to answer RIQ with respect to causal efficacy, they would need to produce a 3x3 table with conditional probabilities of the observed dispersal pattern given conditions involving each of the causal factors that they are investigating: competition for food sources, competition for mates, and inbreeding avoidance.

[7] While there is disagreement about the correct measure of causal strength, the approach of comparing the efficacy of causes with respect to the effect that I defend here is compatible with at least seven measures of causal efficacy (see Fitelson & Hitchcock's ([2011]) survey on these). Determining causal efficacy becomes less clear-cut when the causes under investigation are not dichotomous (Sober [1988]; Wright *et al.* [1993]), but I maintain that there are some cases in which assessing causal efficacy is possible and have suggested one way in which it might be assessed.

[8] Philosophers interested in the first approach to RIQ, the spot-lighting approach, also use terms like 'causal efficacy' and 'bigger difference-makers' in discussing how one should go about selecting certain causes and backgrounding of others. Building on Woodward's ([2003]) interventionist approach to causation, Waters ([2007]) defends a three-way distinction

In cases where there are two or more instances of the explanandum under investigation, each of which is the result of a single cause, a relative frequency question may be asked but not a relative efficacy question (lower-left cell in Table 1). An example mentioned earlier satisfies these conditions—the controversy concerning the importance of allopatric and sympatric speciation in multiple species of fish. This is investigated by estimating how many fish speciation events are allopatric and how many are sympatric. Here 'more important' means 'more frequent'.

The lower-right cell in Table 1 represents instances where both kinds of relative significance controversies occur. Some controversies belonging to this third category ask about the frequency of a cause's efficacy compared to the other causes—Of these instances of the

---

between a 'potential difference making cause,' 'the actual difference making causes,' and 'an actual difference making cause,' where any variable that satisfies the interventionist criterion is a potential difference making cause. This view helps resolve explanatory questions like the one involving the match and the fire. Both the striking of the match and oxygen being present are potential difference making causes because by intervening on either of these, there would be an associated change in explanandum. However, Waters' reason for spot-lighting the striking of the match as 'the' cause of the fire starting rather than another potential difference making cause (oxygen being present) is that the match striking is the variable that actually varied in the situation we are trying to explain; oxygen was around for ages. Waters argues that when variation in the effect variable is associated with actual variation in more than one cause variable, we must instead talk about the set of actual difference making causes (p. 571). Suppose a car crashes because of both an increase in the car's speed and a patch of ice on the road. In this case, there is actual variation in two causal variables that leads to the event we want to explain. This is the situation that researchers investigating the evolution of dispersal patterns are in, but Waters' framework does not provide the tools that researchers need to be able to make the comparative claim regarding the efficacy of an actual difference making cause versus another (that the dispersal pattern is more sensitive to changes in the level of relatedness than changes in the amount of food sources). The view I defend here is closer to Ross's ([forthcoming] p. 9) probable causal control criterion. She presents this criterion in the following way: 'Consider a light switch C, which can take the values "up" or "down" and a light E, which can take the values "on" or "off." In the first case, turning the switch "up" results in a 99% probability of the light bulb being "on," and turning the switch "down" results in a 99% probability of the light bulb being "off." In a second scenario, turning the switch "up" has only a 60% probability of causing the light to turn "on," and turning the light switch "down" only has a 60% probability of turning the light "off." In both cases, the switches have some causal control over the state of the light, but their control differs with regard to how probable each outcome of the contrast is with interventions on the switch'. While this framework provides a way of comparing the efficacy of causes, the comparison is not with respect to an actual situation where two causes simultaneously act to produce an effect. By comparing an actual state of affairs (where both relatedness and availability of food play a role in producing the observed dispersal pattern) with two counterfactual situations (in which one but not the other of relatedness and food are absent), one needs a different analysis of how different causes all affecting a single effect differ in their efficacy. The view I defend here also differs from Woodward's ([2010]) stability criterion, which concerns the extent to which a causal relation will hold when other background conditions change. According to this criterion, if many background conditions must be maintained for a causal relation to hold, then the causal relation is 'unstable'. If, however, the causal relation is preserved even if many of the background conditions are different, then the causal relation is 'stable'. In the hypothetical experiment I propose above, researchers are licensed to conclude that inbreeding avoidance is a stronger cause of the particular dispersal of this particular population; this is a different claim than one about the stability of the causal relationship between inbreeding avoidance and the dispersal pattern.

explanandum, how often is this cause a bigger difference-maker than the other cause(s)? I think this third kind of controversy is more common than one might initially suspect. Earlier, I described the speciation case as a clear example of a relatively frequency controversy and similarly, used the dispersal case as an example of a pure relative efficacy controversy.[9] However, I think both examples are actually better described as this third type of controversy, which I will refer to as a 'blended controversy'.

2.2 Blended controversies

Returning to the contrast between sympatric and allopatric speciation, consider one of the exemplar cases of sympatric speciation: the cichlid species in Lake Victoria. The leading hypothesis for these speciation events is that subpopulations of the fish occupied different microhabitats in which different phenotypes such as eye length and coloration would be adaptive. Thus, there was disruptive selection acting on the cichlid population and there was also non-random (assortative) mating due to the habitat specialization of these subpopulations, which led to the speciation events (McCune and Lovejoy [1998]; Magalhaes *et al.* [2009]).

What may be gleaned from this example is a more nuanced understanding of the causes of speciation events. The kind of geographic separation that leads to reproductive isolation might be an externally imposed barrier in the environment (such as a mountain range), but it might also be a more subtle boundary. In the case of the cichlids, the geographic separation was the result of a selective process (habitat specialization) that isolated subpopulations of fish within Lake Victoria. Here, scientists began asking efficacy questions concerning the strength of the physical barriers that separated subpopulations and the strength of disruptive selection on polymorphic traits that might 'sort' subpopulations into distinct habitats, and in doing so, identified additional speciation

[9] The case of speciation is often used as an example of a pure relative frequency controversy. See (Beatty [1997]; Bird *et al.* [2012]; Kovaka [2017])

mechanisms that were in between ‘pure allopatric’ and ‘pure sympatric’. Indeed, since the shift in this controversy, scientists have come to question the usefulness of the sympatric/allopatric categorization scheme itself (Bird *et al.* [2012]). Dietrich ([2020] p. 248) suggests that this is a general dynamic of relative significance controversies: while the scientists engaging in the controversy tend to start with highly polarized positions, their positions become depolarized over time.

Notice how this description adds nuance to what is at stake in the controversy concerning allopatric and sympatric speciation. When doubt was cast on whether sympatric speciation ever occurred in nature or was merely a theoretical possibility, the theorists involved were inquiring about whether (or how often) selective pressures could be strong enough for subpopulations to evolve and become reproductively isolated from one another. In other words, this controversy began with researchers asking a frequency question, which then shifted to researchers asking about the causal mechanisms that generated sympatric speciation events. This led to another shift: ‘How often are the mechanisms of sympatric speciation causally efficacious?’[10] Although I will delve into this more in Sections 3 and 4, the answer to this question is more substantive than the answer to, ‘how often are speciation events the result of a geographic barrier?’ It also becomes far more understandable why this debate seems meaningful and important to evolutionary biologists who have engaged in it.

The controversy involving the evolution of dispersal patterns also seems better framed as a blend of frequency and efficacy questions although it did not start out that way. In reflecting on previous work on the evolution of dispersal patterns, Dobson and Jones [1985] argue that previous

[10] Dietrich’s (2020) observation that the positions of scientists (often) change in both subtle and more radical ways as the relative significance controversy extends over time is complementary to the view I defend here—that the question at the heart of the controversy can shift.

investigators mistakenly concluded that dispersal patterns must have evolved because of a single cause. These theorists generalized from their data to conclude that, '…the major cause of natal dispersal…among mammals is competition… and the "inbreeding hypothesis" is both inadequate and unnecessary to explain general dispersal patterns' (Moore and Ali [1984] p. 107). This claim suggests that initially, Moore and Ali took the evolution of dispersal patterns to be a relative frequency controversy; an observed dispersal pattern could be explained by one of three causes and to make a general claim about mammalian dispersal patterns, one simply needed to count up the cases. In response, Dobson and Jones argue that this approach is mistaken, as these hypotheses, '…are not exclusive… Even if one cause is of primary importance in some species, other factors may contribute to the magnitude of dispersal…' ([1985] p. 855). Thus, they are interested in efficacy claims. However, Dobson and Jones are still in the business of making general claims about dispersal patterns of classes of organisms, where how often these causes are the biggest difference-maker may be important.

Three points should be noted here. First, observe that blended controversies may result in 'conflicting' answers to RIQs; while C_1 might be stronger than C_2 in a few cases, the more frequent arrangement may be that C_2 is stronger than C_1. However, the 'conflict' is only apparent because two different senses of 'causal importance' are in play. Although scientists may disagree on which relative significance controversy is more pertinent to their investigation, a set of conflicting answers need not mean that the scientists are disagreeing about the phenomenon of interest.[11] Second, observe that blended controversies allow for scientists to still talk about the more

[11] The spot-lighting approach that I highlighted in the introduction introduces another way in which claims about relative importance may come into apparent conflict, without always being in actual tension. Using the conception of causal efficacy I described above, oxygen being present and the match being struck are equally efficacious, but one cause may be regarded as more important than the other with respect to the spot-lighting and selecting of causes approach to relative importance.

efficacious cause across two or more instances of the explanandum where C_1 is more efficacious in some cases while C_2 is more efficacious in others. Finally, earlier analyses of relative significance controversies have not been sufficiently appreciative of how these two kinds of controversies can co-occur (the blended format). This is important because if these controversies are ubiquitous (as I suggest they are), ignoring this third category may hinder progress on the question concerning the epistemic value of engaging in these controversies.

2.3 Reference classes in relative significance controversies

There is one more element to add to our analysis of relative significance controversies. So far, we have been tacitly assuming a well-defined reference class for each controversy—e.g., 'all fish speciation events' and 'prairie dog dispersal patterns' in the second. But this is a further feature of relative significance controversies—defining the instances of the explanandum that are of interest. When there are two or more instances of the explanandum that one is interested in, the relative significance controversy is married to a specified reference class. Without a reference class, relative significance questions are incompletely specified. Thus, in addressing the controversy, one provides a statement assessing the relative importance of the cause(s) with respect to a reference class. These reference classes are determined by the goals of the scientists who initiate the controversy and are essential to pinpointing the relative significance controversy that one is engaged in.

The example of speciation suggests that if one asks how often speciation events for all living things are allopatric as opposed to sympatric, the former will have a much higher frequency than the latter. Suppose, however, that the reference class is changed to 'fish species living in lakes where there are many closely related fish species'. In this case, sympatric speciation has a significantly higher frequency than it does in the first case (McCune and Lovejoy [1998]). To say that allopatric speciation is 'more important' in the first case does not entail that it will also be in the second. The apparent disagreements that the scientists engaging in this controversy are having may not be about the shared phenomenon of interest but instead, may lie in the fact that they are asking different questions or the same question but with respect to different reference classes.

Using the analysis I offer, if we are to understand what a relative significance controversy is about (and how it will be settled), we need to understand two things: the reference class it pertains to, and the kind of controversy that it is (relative frequency, relative efficacy, or a blend of both). I next turn to the doubts one may have concerning the epistemic worth of these controversies.

3. The Epistemic Challenge to Relative Significance Controversies

Philosophical analyses of relative significance controversies arose in conjunction with John Beatty's [1995] influential Evolutionary Contingency Thesis (ECT). The ECT asserts that there are no distinctively biological laws, by which Beatty means generalizations that are empirical, necessary, and biological as opposed to physical or chemical, and are nomologically necessary. Beatty argues biological generalizations hold (to the extent that they do) as a result of contingent evolutionary forces—combinations of mutation, natural selection, drift, and other probabilistic processes. In virtue of these contingencies, it will often be the case that several single-factor theories are each responsible for some proportion of instances of the phenomenon of interest.[12] It is, therefore, natural for scientists to then ask what these proportions are. This commits biologists to theoretical pluralism, which sets the stage for relative frequency controversies to arise.[13]

[12] See (Beatty [1987]; Dietrich [2020]) for an insightful analysis of the classical/balance controversy. An example of this kind of relative significance controversy used in (Beatty [1997]).

[13] The multicausal processes that Beatty is interested in are of the first type I describe where a single type of explanandum is sometimes caused exclusively by one cause and sometimes exclusively by another (e.g., allopatric vs. sympatric speciation). Some have questioned whether theoretical pluralism follows from the ECT and whether relative significance controversies are forced on biologists in the absence of biological laws. I will not address these questions here, but for more on this, see Sober [1997].

Although Beatty argues that the absence of biological laws is what gives rise to relative significance controversies, I suggest these controversies arise even if there are biological laws.

3.1 Is engaging in a relative frequency controversy valuable?

According to Beatty, the ECT explains why relative frequency controversies arise frequently in evolutionary biology, but he argues there is an unanswered normative question about whether these controversies are worth pursuing. He takes up the following question concerning the value of frequency controversies: *What does one gain by resolving a relative frequency controversy?* The answer to a relative frequency question is a tallying statement of the form: X_1% of the instances of Y are caused by Z_1, while X_2% of the instances of Y are caused by Z_2, etc... These frequency distributions are subject to evolutionary forces, and so resolving a controversy at one time does not mean the controversy will remain resolved. Because of this, Beatty ([1997] pp. S438-40) thinks such tallying statements are not 'interesting,' 'telling,' or representative of anything 'more than a highly contingent fact of evolutionary history'.[14]

Kovaka ([2021] p. 7788) develops this criticism further, contrasting the relative frequency distribution of a biological phenomenon with the relative frequency distributions of phenomena that are determined by physical laws. She states:

> There *is* [emphasis added] something telling about the 51% probability that when you flip a coin it will come up the same way it started [Diaconis *et al*. 2007). In the coin case,

[14] It should be noted that Beatty's argument is ambiguous; it is unclear which of the following candidate theses he is defending: (i) Relative frequency controversies <u>may be</u> uninteresting, (ii) Relative frequency controversies <u>are sometimes</u> uninteresting, or (iii) Relative frequency controversies <u>are always</u> uninteresting. The third candidate, 'Relative frequency controversies are always uninteresting,' seems to be closest to his position since (i) is trivially true, and in his analysis, Beatty considers two collectively exhaustive subsets of relative frequency controversies: those where the difference in the distributive frequency of the outcome brought about by various causes is small (e.g., 45%, 55%) and those where it is large (90%, 10%). In both cases, he suggests that these controversies are pursued for socioeconomic reasons rather than epistemic ones (Beatty [1997] pp. S440-41). This lack of specificity also obscures the comments he makes here: if Beatty equates 'uninteresting' with 'not being more than a highly contingent fact of evolutionary history,' this implies that all biological research is uninteresting, which he surely does not believe.

> physical laws determine the relative frequency of heads to tails, but in the adaptation vs. drift case, there is no deeper principle that explains or determines relative frequency. This point is underscored by the fact that biological relative frequency distributions themselves can evolve and change, so having resolved a relative frequency controversy at one time does not mean that the answer was always or will continue to be correct.

Unlike the case of the coin flips where Kovaka suggests there is a deeper principle at work, she takes Beatty to view the speciation frequency distribution as an enumerative fact-gathering endeavour that is neither about a scientifically interesting question nor will it yield anything epistemically valuable.[15]

3.2 A direct source of epistemic value

Beatty's challenge has two components: (i) What is the point of settling relative frequency controversies, given that frequency data are subject to evolutionary forces and will probably change? and (ii) Why is a tallying statement about frequencies ever scientifically interesting or epistemically valuable? One initial answer to (i) is that evolutionary biology *is* backward-looking and in trying to explain an evolutionary claim like, 'why do we see the traits that we presently do?', we need frequency data concerning the past, even if the future will be different. In this section, I offer a second response to (i), but before doing so, I want to further outline (ii). Given his scepticism about what is gained by learning a tallying statement about frequencies, Beatty indicates that merely 'increasing knowledge' is not a sufficient condition for being epistemically

[15] Diaconis *et al.* [2007] note that an inhomogeneity in the mass distribution of the coin would change the outcome of coin flips. Therefore, it is a contingent fact about the way the coin is made that determines this probability, not a deeper principle. What is more, the fact that Pr(Coin lands heads | Coin is tossed) = 0.51 depends on the frequency distribution of the upper-momentum and the angular momentum of the toss, which reflect contingent facts about the coin-toss system (Diaconis *et al.* [2007] p. 220).

valuable.[16] In order to fully address (ii) I will interpret Beatty to be asking, 'why is a tallying statement of this kind good for a scientist to know?'[17]

Although Kovaka ([2021] p. 13) does not take a stand on whether relative frequency controversies are in and of themselves interesting, she does think they are *indirectly epistemically valuable* since these controversies can improve upon existing explanations. Appealing to Angela Potochnik's understanding of explanation, Kovaka maintains that relative frequency controversies further the explanatory project by elucidating an explanation's scope of applicability. Explanations should not only identify causal relationships but the conditions that are necessary for the phenomenon of interest to occur (Potochnik [2017] p. 139). In cases where more than one cause is responsible for a phenomenon of interest, it is valuable to know that X may cause Y but also valuable to know under what conditions X causes Y.

This can be illustrated using the allopatric versus sympatric speciation controversy. It is valuable for scientists to know geographic barriers that divide populations into two or more subpopulations are not necessary for speciation events to occur. But it is also valuable for scientists to know what kinds of conditions will prevent or promote the cause from obtaining. Learning that sympatric speciation events are more likely to occur when closely related species coexist in an isolated environment is an important contribution to our explanation of speciation events. This can be generalized: when scientists engage in one of these controversies, they are positioned to uncover properties that link one of the postulated causal mechanisms to a subset of explananda events.

This discovery is important because scientists infer something about the causal structure of the phenomenon of interest—namely, how causes cluster—and adopt a reference class that

[16] Many thanks to an anonymous reviewer for asking me to clarify this point.

[17] I grant that this question can only be answered on a case-by-case basis, however, in what follows, I show why a general answer—particularly Beatty's sceptical one—is unsatisfactory.

corresponds to these clusters. Thus, relative frequency controversies are indirectly valuable because they are a means to improving our understanding of the explanatory scope of the phenomenon.[18] While I agree that relative frequency controversies can improve our understanding of explanatory scopes indirectly, I think engaging in relative frequency controversies can also be directly epistemically valuable.

The predictively powerful statements that scientists learn by resolving these controversies are one direct source of epistemic value gained through relative significance controversies.[19] For example, suppose a scientist adopts the following empirical adaptationist's thesis: selection is the primary cause of (and the only cause one needs to postulate to explain) the majority of traits that organisms now have (Orzack and Sober [1994]). Suppose further that this scientist then encounters a new trait. What she gains from adopting this thesis is an expectation of what kind of explanation is usually correct and what kind of explanation is not. These findings may be used to assign prior probabilities about the evolutionary history of newly discovered traits. When she encounters the new trait, therefore, her prior probability that the presence of the trait is the result of selection as opposed to another evolutionary cause should be pretty high. In other words, a frequency distribution of the causes for some set of explananda events provides scientists with a prior-setting mechanism that assists them in forming predictions about similar explananda. Further, even if Beatty is right that this relative frequency is contingent, knowing the relative frequency of outcomes brought about by the set of causes as the world is now can help scientists determine what kinds of explanations for a phenomenon might be correct. In this way, relative frequency distributions are informative and act as guides to future research.

[18] Notice that if there were another way of uncovering the explanatory scope, Beatty's question, again, arises: Why should scientists engage in relative frequency controversies?
[19] This source of epistemic value provides an answer to both (i) and (ii) outlined at the beginning of Section 3.1.

This point carries over to less extreme cases. Beatty ([1997] p. S438) considers a controversy with a seemingly small difference in relative frequency: two causes, C_1 and C_2, are thought to explain 55% and 45% of the instances of a phenomenon respectively, but suppose a scientist discovers that C_1 explains 45% of the instances, and C_2 explains 55% of them. Beatty takes this case to be an obvious example of a relative frequency controversy that is not worth engaging in. I disagree; while this may not be as dramatic a finding as other frequency distribution discoveries may be, the 45%, 55% tallying statement is still epistemically valuable because it influences the predictions the scientist will make about other instances of this phenomenon. In short, both modest and dramatic revisions to relative frequency claims are epistemically valuable.

This understanding of a prediction's epistemic value makes sense of Beatty's claim that a controversy of the 55%, 45% sort is uninteresting, whereas the controversy associated with the empirical adaptationist thesis is very interesting. However, recall the focus on evolutionary biology in Beatty's objection: relative frequency statements that concern evolutionary causes are contingent and subject to evolution. If this is true, why should these relative frequency claims guide future research? This response is only successful if scientists' priors do not continue to update in response to changes in frequency claims.

In this way, the epistemic value of learning the tallying statement about frequency is gradable: the epistemic value of learning the statement corresponds to how big the numerical change in the probability is before and after updating. This shows how engaging in a relative frequency controversy can be directly epistemically valuable. In this case, the controversy is not a means to something further, but the tallying statement about the causes can, itself, be epistemically valuable. However, it is not only the predictive power or changes in reference class that make relative significance controversies well worth engaging in.

4. Relating Relative Frequency Controversies to Other Significance Controversies

Up to this point, I have pushed to expand Kovaka's qualified defence of the epistemic value of relative frequency controversies by arguing that such controversies, considered in their own right, have both indirect and direct epistemic value. I now turn to another way in which these controversies can be indirectly epistemically valuable. Beatty and Kovaka are careful to distinguish relative frequency controversies from other significance controversies. Beatty ([1997] p. S433) acknowledges that relative efficacy controversies occur in biology, however, his comments are directed towards relative frequency controversies specifically, which he takes to be 'a different family of issues'. Kovaka ([2017] p. 101) acknowledges that other kinds of controversies exist,[20] but argues that relative frequency controversies are not just conceptually distinct from other relative significance controversies but that 'the structure of disagreements about relative frequency and relative significance is also importantly different'.

While I agree that relative frequency controversies are a unique kind of significance controversy, I do not think their epistemic value can be captured by looking at frequency controversies in isolation from other sorts of significance questions that are simultaneously being investigated or that the frequency controversy gives rise to. As I illustrated in Section 2.2, relative significance controversies in evolutionary biology that may appear to be controversies about frequency are better described as blended controversies. Is this true generally? I think so. When scientists inquire about which causes are more important for producing some instances of a

[20] For example, in discussing a controversy concerning the relative significance of drift and selection in the observed variation of beak length in Galapagos finches, Kovaka ([2021] p. 7780) highlights how these controversies can be about efficacy. She states: 'A combination of selection and drift can produce evolutionary change in the same trait, which makes disentangling their relative contributions to particular instances exceptionally difficult'.

phenomenon, this question could be interpreted as either a question about frequency or efficacy. In either case, inquiring about one of these interpretations of ‘relative importance’ will lead naturally to the other. For example, after learning that a cause has a high frequency, a reasonable next step is to explore whether (and how often) the cause is a difference-maker. The blended framework will better describe the controversies where both interpretations are explored. Furthermore, it is only in light of this analysis, I’ll argue in the rest of Section 4, that we can develop a more complete account of the epistemic value of frequency controversies. When relative frequency controversies are blended with relative efficacy controversies, the two kinds of controversies inform each other—with the upshot that frequency claims can act as a check on efficacy claims (and vice versa). The rest of this section is dedicated to demonstrating this by looking at two examples from the history of evolutionary biology.

4.1 The Neutralist-Selectionist controversy

The neutralist-selectionist controversy began when studies revealed the number of amino acid substitutions between two species was approximately proportional to the time divergence of the two species. Using this data, Kimura ([1968]) found that the rate of nucleotide substitutions was enormously higher than what had been previously estimated and could not be explained by natural selection alone; these substitutions were occurring much more rapidly than what would be predicted if selection were the ‘driving force’ of molecular evolution. From this data, Kimura suggested that most of these substitutions were not deleterious or beneficial but neutral and could drift to a high frequency or even fixation. This sparked a relative frequency controversy concerning the percentages of substitutions that were non-neutral versus neutral. Kimura’s theory was in tension with selectionism, which assumed that mutations that reached fixation were almost always beneficial. Observe that here, the relative frequency controversy originated with an inconsistency between data and theory in molecular evolutionary biology.

The controversy ensued and molecular biologists used comparative sequence analyses to estimate the proportion of mutations that went to fixation as a result of selection. Quickly, evidence emerged that seemed to contradict Kimura's predictions. To resolve this tension, Ohta ([1973]) suggested that a distinction be drawn between neutral and nearly neutral or 'mutants belonging to the borderline class,' which would account for the discrepancies between observations and Kimura's predictions (Ohta [1973] p. 270). Unlike strictly neutral mutations, Ohta argued mutations that were only slightly deleterious would be weakly selected against, and so could find their way to fixation. By the same token, genes that are only slightly advantageous can fail to go to fixation when they evolve in small populations. Ohta ([1973]) hypothesized that nearly neutral mutations would act as if they were neutral and would often drift to a high frequency and then to fixation.

In this case, what scientists learned from Kimura's and later, Ohta's work, were facts about the causal efficacy of variation in fitness. These nearly neutral mutations were subject to both selection and drift, and the question of interest became: 'Of these nearly neutral mutations, how often does drift "dominate" selection?'[21] In other words, what started out as a simple relative frequency controversy became a blended controversy as these theorists began to understand more about the causal structure of molecular evolution. This example illustrates how a relative frequency controversy can inform scientists about the relative efficacy of the causes that they are invoking.[22] By making headway on the frequency question, researchers realized that the rates of fixation of neutral mutations were inconsistent with Kimura's initial prediction and needed explanation. This was provided by Ohta, who showed there was a relation between the relative efficacy of selection

[21] See (Gillespie [1999]; Dietrich and Millstein [2008]; Clatterbuck, Sober, and Lewontin [2013]) for different characterizations of the question about the neutralist-selectionist controversy and the causal importance of drift and selection.

and drift, which could be observed by adjusting the difference in the fitness of alleles (a ratio measuring the strength of selection) and the given population's size.

The transition from a relative frequency controversy to a blended controversy that occurred in the neutralist-selectionist debate is no coincidence. As I noted earlier, the controversy started because of an inconsistency between what theorists believed to be the more efficacious cause (selection) and the available data. Relative frequency controversies are a tool that allows scientists to inquire about the efficacy of causes. Because of this, we will often see an efficacy controversy lurking just below a relative frequency controversy.

4.2 The controversies surrounding group selection

There are two approaches scientists may take to investigate phenomena where more than one causal relationship has been identified—a top-down and a bottom-up approach. One uses the top-down approach by uncovering theoretical reasons for why a cause should be rare or common (or more/less efficacious). Alternatively, one uses the bottom-up approach by inferring from observations which causes are common and which are rare (or which are more/less efficacious).

G. C. Williams's [1966] analysis of group selection provides an example of the top-down approach. While he thinks evolution by group selection is a theoretical possibility, he is sceptical that it ever makes a difference in evolutionary outcomes (Williams [1966] p. 122-23) and formulates several arguments to defend this point. One of them states that in order for group selection to occur, one of the following conditions must be met: either the species in question has a particular population structure in which the groups are small and migration is minimal or the trait selected for at the group-level is not disadvantageous at the individual-level (Williams [1966]).[23] Because each of these conditions is rarely satisfied (and even when one of them is, there is no

[23] While group selection may occur when this latter condition holds, Williams argues, given parsimony considerations, that it ought not be invoked if individual selection alone can explain the evolutionary change (Williams [1966] p. 19).

guarantee that the condition will persist long enough for group selection to cause a trait to evolve), Williams argues that group selection is generally much weaker than individual selection and that these constraints provide ‘a reasonable basis for scepticism about the effectiveness of this evolutionary force’ (Williams [1966] p. 116). Notice that here, Williams uses theoretical justification about group selection’s efficacy to support a prediction about how often group selection makes an important contribution to evolutionary change; he is using the top-down approach.

Contrast this with Michael Wade’s [1978] summary of empirical research on group selection. Wade’s ([1976]; [1977]) experimental work on group selection sheds light on whether (and when) group selection may ‘override’ individual selection. In his [1978] paper, Wade uses the data collected from populations of *Tribolium* to suggest that Williams’s model was incomplete and that when his model’s assumptions are modified in biologically realistic ways, the models will be ‘made more favourable to the process of group selection’ (Wade [1978] p. 112). Building on this work, Charles Goodnight and Lori Stevens [1997] point out that Williams considers only additive genetic effects. Using field studies on group selection that had been conducted under natural conditions, Goodnight and Stevens ([1997] p. S65) show that nonadditive effects were the ‘major source of the unique genetic response to group selection’. When nonadditive genetic effects increase, group-level selection is more likely to have a significant effect and result in novel evolutionary outcomes (Goodnight and Stevens [1997] p. S76). These effects, which had not been considered in Williams’ analysis, explained how gene interaction at the organismic level contributes significantly to differences at the group level.[24]

[24] Philosophers have contributed significantly to the literature on what group selection is, as well as the question concerning the strength of group selection compared to individual selection (once group selection has been defined). Although I can’t do service to the depth of the philosophical debate here, see (Lloyd [1988]; Brandon [1990]) for more

Wade as well as Charles and Goodnight use the bottom-up approach to respond to Williams; their argument rests on collected data. Because these studies supported the hypothesis that group selection was the more efficacious cause and assuming that these studies were sufficiently representative, Goodnight and Stevens conclude that their results should be extended to other cases. Here, Goodnight and Stevens respond to the relative frequency claim that Williams postulated by arguing that group selection does appear to contribute to evolutionary outcomes more often than previously believed.

This example from the group selection debate illustrates the relation between relative frequency and relative efficacy controversies: the efficacy of a cause predicts the frequency of its effects. If group selection is a weak force, then we should expect to see very few evolutionary outcomes that are the result of group selection. However, when empirical data is acquired that appears inconsistent with those predictions, further theoretical work must be done. In this case, Wade [1978] as well as Goodnight and Stevens [1997] uncovered important features of the causal relations they were studying that could undermine previous efficacy claims.

We have now seen how frequency claims can act as a check on efficacy claims (and efficacy claims can act as a check on frequency claims), and why these historical examples, which appear to be relative frequency controversies, are better understood as controversies of the blended form. In the neutralist-selectionist controversy, a frequency claim drove researchers to investigate the respective strengths of the underlying causes of molecular evolution—drift and selection—and in doing so, disconfirmed a previous hypothesis about how often drift was the most efficacious cause. In the group selection case, a hypothesis about the relative efficacy of group selection was tested and in conflict with acquired frequency data. This led researchers to revise their

on the first question, and see (Sober and Wilson [1998]; Griesemer and Wade [1988]; Okasha [2006]) for more on the second.

understanding of why group-level selection was rarely a more efficacious cause than individual-level selection.

As I suggested earlier, I think the blended form of relative significance controversies is ubiquitous in evolutionary biology, but they often start as—or appear to be—controversies that are purely about frequency. When these controversies do take the blended form, engaging in them can have singular epistemic upshots. An interesting question that deserves to be investigated in future work on blended controversies is whether there is a more general pattern of frequency claims acting as 'checks' on efficacy claims (and vice versa) throughout evolutionary biology, or if this is a unique feature of biologically important debates like the neutralist-selectionist and group selection controversies that I consider here. The relationship between frequency and efficacy is not straightforward: as I have shown, causes can have high frequency but not be the most efficacious cause and they can be rare and yet highly efficacious when they occur. If it turns out that this epistemic upshot is not often produced by engaging in blended controversies, this is compatible with my claim that the framework of blended controversies often allows for a richer and more nuanced description of relative significance controversies. However, a hint of Beatty's scepticism about the value of engaging in these controversies may be warranted, if these controversies rarely produce the epistemic upshot I have described here.[25]

5. Conclusion

Engaging in relative frequency controversies is an epistemically valuable scientific endeavour for at least three reasons. First, these controversies help scientists form predictions about new instances of the phenomenon of interest. Second, they refine our understanding of the causes of the phenomenon of interest and can help in identifying an appropriate reference class that fixes the

---

[25] Many thanks to an anonymous reviewer for asking me to clarify this point.

explanatory scope of hypotheses. Finally, many relative frequency controversies are better described as blended controversies; while some controversies—like the neutralist-selectionist and group selection controversies—seem to be about frequencies, they also involve claims about efficacy. Because of this relationship, claims about frequency can inform scientists about efficacy (and vice versa). In these cases, relative frequency controversies can also improve our understanding of causal systems in evolutionary biology.

Acknowledgments

I am very grateful to Elliott Sober and Hayley Clatterbuck as well as Rachael Brown, Stephanie Hoffmann, Matt Maxwell, Greg Nirshberg, Mike Titelbaum, Shimin Zhao, and four anonymous reviewers for their constructive comments on earlier drafts of this manuscript. I also thank the audiences at the 2021 APA-Pacific, 2021 CSHPS meeting, 2021 NM-TX Philosophical Society's annual meeting, and 2021 ISHPSSB meeting for their encouragement and helpful suggestions.

Department of Philosophy
University of Wisconsin-Madison
5185 Helen C. White Hall, 600 N Park St., Madison WI 53706, USA
Email: kdeaven@wisc.edu

References

Beatty, J. [1987]: 'Weighing the Risks: Stalemate in the Classical/Balance Controversy', *Journal of the History of Biology*, **20**, pp. 289–319.

——— [1995]: 'The Evolutionary Contingency Thesis', in J. Lennox and G. Wolters (*eds*), *Concepts, Theories and Rationality in the Biological Sciences*, Pittsburgh: University of Pittsburgh Press, pp. 45–81.

——— [1997]: 'Why Do Biologists Argue Like They Do?', *Philosophy of Science*, **64**, pp. S432–43.

Bird, C. E., Fernandez-Silva, I., Skillings, D. J. and Toonen, R. J. [2012]: 'Sympatric Speciation in the Post "Modern Synthesis" Era of Evolutionary Biology', *Evolutionary Biology*, **39**, pp. 158–80.

Brandon, R. N. [1990]: 'Adaptation and Environment', Princeton: Princeton University Press, <https://press.princeton.edu/books/hardcover/9780691630496/adaptation-and-environment>.

Diaconis, P., Holmes, S. and Montgomery, R. [2007]: 'Dynamical Bias in the Coin Toss', *SIAM Review*, **49**, pp. 211–35.

Dietrich, M. R. and Millstein, R. L. [2008]: 'The Role of Causal Processes in the Neutral and Nearly Neutral Theories', *Philosophy of Science*, **75**, pp. 548–59.

Dietrich, M. R. [2020]: 'What Is the Nature of Scientific Controversies in the Biological Sciences?', in K. Kampourakis and T. Uller (*eds*), *Philosophy of Science for Biologists*, Cambridge: Cambridge University Press, pp. 235–54, <https://www.cambridge.org/core/books/philosophy-of-science-for-biologists/what-is-the-nature-of-scientific-controversies-in-the-biological-sciences/A4436E73960BA79A30E21B64C12419B2>.

Endler, J. A. [1986]: 'Natural Selection in the Wild', Princeton University Press.

Fitelson, B. and Hitchcock, C. [2011]: '29 Probabilistic Measures of Causal Strength', in P. McKay Illari, F. Russo and J. Williamson (*eds*), *Causality in the Sciences*, Oxford University Press, p. 0, <https://doi.org/10.1093/acprof:oso/9780199574131.003.0029>.

Franklin-Hall, L. R. [2015]: 'Explaining Causal Selection with Explanatory Causal Economy: Biology and Beyond', in P.-A. Braillard and C. Malaterre (*eds*), *Explanation in Biology: An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*, Dordrecht: Springer, pp. 413–38.

Gillespie, J. H. [1999]: 'The Role of Population Size in Molecular Evolution', *Theoretical Population Biology*, **55**, pp. 145–56.

Godfrey-Smith, P. [2001]: 'Three Kinds of Adaptationism', in S. H. Orzack and E. Sober (*eds*), *Adaptationism and Optimality*, Cambridge: Cambridge University Press, pp. 335–57.

Goodnight, C. J. and Stevens, L. [1997]: 'Experimental Studies of Group Selection: What Do They Tell US About Group Selection in Nature?', *The American Naturalist*, **150**, pp. s59–79.

Griesemer, J. R. and Wade, M. J. [1988]: 'Laboratory Models, Causal Explanation and Group Selection', *Biology and Philosophy*, **3**, pp. 67–96.

Hughes, A. L. [2008]: 'Near Neutrality', *Annals of the New York Academy of Sciences*, **1133**, pp. 162–79.

Kimura, M. [1968]: 'Evolutionary Rate at the Molecular Level', *Nature*, **217**, pp. 624–6.

Kovaka, K. [2017]: 'Understanding Innovation And Imitation In Evolution', University of Pennsylvania.

——— [2019]: 'Underdetermination and Evidence in the Developmental Plasticity Debate', <https://philpapers.org/rec/KOVUAE>.

——— [2021]: 'Fighting about Frequency', *Synthese*, **199**, pp. 7777–97.

Lewis, D. [1986]: 'Causal Explanation', in D. Lewis (*ed.*), *Philosophical Papers*, Vol. II New York: Oxford University Press, pp. 214–40.
Lewontin, R. C. [2006]: 'The Analysis of Variance and the Analysis of Causes', *International Journal of Epidemiology*, **35**, pp. 520–5.
Lloyd, E. A. [1994]: 'The Structure and Confirmation of Evolutionary Theory', Princeton: Princeton University Press, <https://press.princeton.edu/books/paperback/9780691000466/the-structure-and-confirmation-of-evolutionary-theory>.
McCune, A. R. and Lovejoy, N. R. [1998]: 'The Relative Rate of Sympatric and Allopatric Speciation in Fishes: Tests Using DNA Sequence Divergence between Sister Species and among Clades', in *Endless Forms: Species and Speciation*, Oxford: Oxford University Press, pp. 172–85.
Mill, J. S. [1874]: 'A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation', 8th ed., New York: Harper & Brothers.
Moore, J. and Ali, R. [1984]: 'Are Dispersal and Inbreeding Avoidance Related?', *Animal Behaviour*, **32**, pp. 94–112.
Ohta, T. [1973]: 'Slightly Deleterious Mutant Substitutions in Evolution', *Nature*, **246**, pp. 96–8.
Okasha, S. [2009]: 'Evolution and the Levels of Selection', Oxford, New York: Oxford University Press.
Orzack, S. H. and Sober, E. [1994]: 'Optimality Models and the Test of Adaptationism', *The American Naturalist*, **143**, pp. 361–80.
Potochnik, A. [2020]: 'Idealization and the Aims of Science', University of Chicago Press.
Ross, L. N. [2021]: '"Causal Control: A Rationale for Causal Selection."', in C. K. Waters and J. Woodward (*eds*), *Philosophical Perspectives on Causal Reasoning in Biology*, Vol. XIX Minneapolis: University of Minnesota Press.
Smith, J. M. [1966]: 'Sympatric Speciation', *The American Naturalist*, **100**, pp. 637–50.
Sober, E. [1988]: 'Apportioning Causal Responsibility', <https://philpapers.org/rec/SOBACR-4>.
Sober, E., Wright, E. O. and Levine, A. [1992]: 'Reconstructing Marxism', *London: Verso*, **59**, p. 133.
Sober, E. [1997]: 'Two Outbreaks of Lawlessness in Recent Philosophy of Biology', *Philosophy of Science*, **64**, pp. S458–67.
Sober, E. and Shapiro, L. [2007]: 'Epiphenomenalism: The Dos and the Don'ts', in P. Machamer and G. Wolters (*eds*), *Thinking About Causes: From Greek Philosophy to Modern Physics*, Pittsburgh: University of Pittsburgh Press, pp. 235–64.
Sober, E. and Wilson, D. S. [n.d.]: 'Unto Others', Cambridge, Mass.: Harvard University Press, <https://www.hup.harvard.edu/catalog.php?isbn=9780674930476>.
Tregenza, T. and Butlin, R. K. [1999]: 'Speciation without Isolation', *Nature*, **400**, pp. 311–2.
Wade, M. J. [1976]: 'Group Selections among Laboratory Populations of Tribolium.', *Proceedings of the National Academy of Sciences*, **73**, pp. 4604–7.
Wade, M. J. [1977]: 'An Experimental Study of Group Selection', *Evolution*, **31**, pp. 134–53.
——— [1978]: 'A Critical Review of the Models of Group Selection', *The Quarterly Review of Biology*, **53**, pp. 101–14.
Walsh, D. M., Lewens, T. and Ariew, A. [2002]: 'The Trials of Life: Natural Selection and Random Drift', *Philosophy of Science*, **69**, pp. 452–73.
Waters, C. K. [2007]: 'Causes That Make a Difference', *The Journal of Philosophy*, **104**, pp. 551–79.
Williams, G. C. [1966]: 'Adaptation and Natural Selection Princeton University Press', *Princeton, NJ*,.

Woodward, J. [2005]: ‘Making Things Happen: A Theory of Causal Explanation’, Oxford university press.
——— [2010]: ‘Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation’, *Biology & Philosophy*, **25**, pp. 287–318.
——— [2021]: ‘Causation with a Human Face: Normative Theory and Descriptive Psychology’, Oxford University Press.