

A Priori Causal Models of Natural Selection

Elliott Sober

Abstract: To evaluate Hume's thesis that causal claims are always empirical, I consider three kinds of causal statement: " e_1 caused e_2 ," " e_1 promoted e_2 ," and " e_1 would promote e_2 ." Restricting my attention to cases in which " e_1 occurred" and " e_2 occurred" are both empirical, I argue that Hume was right about the first two, but wrong about the third. Standard causal models of natural selection that have this third form are *a priori* mathematical truths. Some are obvious, others less so. Empirical work on natural selection takes the form of defending causal claims of the first two types. I provide biological examples that illustrate differences among these three kinds of causal claim.

Keywords: *a priori, causation, Davidson, evolution, Hume, laws of nature, natural selection.*

1. Introduction

Hume was wrong about what causality is, but was he right in claiming that causal statements are always empirical? Favorable examples readily come to mind. Consider the statement that smoking causes lung cancer. Reason alone does not suffice to show that this statement is true; observations are necessary. Examples like this make Hume's thesis seem obvious. But Hume had something more than favorable examples to support his general claim. He had an argument. In Book I, Section IV, Part 1 of the *Enquiry Concerning Human Understanding*, Hume says that "... every effect is a distinct event from its cause. It could not, therefore, be discovered in the cause." This argument begins

with metaphysics and ends with epistemology (call it “the M/E argument”). Does the epistemology follow from the metaphysics?

There is another argument for Hume’s thesis. Statements of the form “event e_1 caused event e_2 ” entail that e_1 occurred and that e_2 did too. Can the occurrence of an event be known *a priori*? If not, causal statements with the form just described must be empirical. I don’t propose to defend the claim that statements of the form “event e occurred” are always empirical. Perhaps the claim has counterexamples — e.g., Descartes’ *cogito*. The proposition that I (now) am thinking surely describes an event, one that has causes and effects. But is it *a priori*? Intuitively, the *a priori* grasp of a proposition (through “reason alone”) differs from the introspective grasp of one’s own mental state. Unfortunately, explaining why this is so is not so straightforward. But, lucky for me, the goal of this paper does not oblige me to face this issue. My goal is to discuss causality and the *a priori* in evolutionary theory and in scientific theories more generally. The epistemic status of 1st person reports is not germane to that topic, since scientific theories can be represented adequately via purely 3rd person statements. And when the sciences make claims of the form “event e_1 caused event e_2 ,” it is pretty much always true that “event e_1 occurred” and “event e_2 occurred” are both empirical. So I will restrict my claim about “event e_1 caused event e_2 ” to the following conditional: if “ e_1 occurred” or “ e_2 occurred” are empirical, the causal statement must be empirical as well.

This argument shows what is wrong with Davidson’s (1963, p. 696) claim to have found a counterexample to Hume’s thesis. Davidson says that statements of the form “the cause of e_2 caused e_2 ” are analytic, which I assume entails that they are *a priori*. The quoted sentence entails that e_2 occurred and that the cause of e_2 did too. These aren’t

a priori, so the quoted sentence isn't *a priori*, either. Davidson's challenge to Hume's thesis might be successful if "event e_1 caused event e_2 " meant that if an e_1 -ish event occurs, then an e_2 -ish event also occurs. The conditional "if the cause of e_2 occurs, then e_2 occurs" is true *a priori*. The trouble for Davidson derives from the fact that the causal statement and this conditional are not equivalent. This is not to gainsay Davidson's important observation that "the truth of a causal statement depends on *what* events are described; its status as analytic or synthetic depends on *how* the events are described."

Now let's consider a second kind of causal statement. Instead of focusing on *causing* an event, let's consider what is involved in *promoting* its occurrence. Consider "smoking promoted cancer." This statement does not entail that anyone ever got cancer; it just says that smoking put individuals at greater risk.¹ However, the statement does entail that there was some smoking.² Let me emphasize this point by inserting the word "actually." If smoking *actually* promoted cancer, some individuals must have smoked. This shows that statements of the form "event e_1 actually promoted event e_2 " are not *a priori* (assuming, as before, that " e_1 occurred" is not *a priori*).

There is a third causal locution. It involves *potential* promoting, not *actual* promoting. An example is the statement that smoking would promote cancer. This statement does not entail that anyone smoked or that anyone got cancer, so the argument that showed that two other kinds of causal statement are empirical does not deliver that

¹ Causal promotion does not require determinism. Smoking can promote lung cancer even when complete sets of causal conditions always confer intermediate probabilities on the effect term of interest. I think the same is true of "event e_1 caused event e_2 " (the statement does not require determinism), but my arguments in this paper do not require that this is so.

² The sentence "smoking promotes lung cancer" is ambiguous; it could mean that smoking *actually* is promoting lung cancer in some individuals now, or that smoking *would* promote it. Oddly enough, the past tense removes the ambiguity; "smoking promoted lung cancer" means *actually* promoted.

verdict here. But surely, in this example, we have an empirical claim; reason alone does not reveal that smoking would promote cancer. Is there an exception to this pattern?

There is, and it is an old saw. In his play *Le Malade Imaginaire*, Molière makes fun of a philosopher who says that opium puts people to sleep because it has a dormitive virtue. The point of the joke is that the explanation is empty. However, the emptiness of the explanation does not settle whether the relevant causal claim is true. I think it's not just *true* but *a priori* true that ingesting something soporific would promote falling asleep. The fact that the statement is trivial does not show that it is false. The causal statement expresses the truism that things that have a dormitive virtue have a certain type of causal power.

The fact that some claims of the form “*x* would promote *y*” are *a priori* throws light on Hume's M/E argument. Here is another statement of the argument, also from Book I, Section IV, Part 1 of the *Enquiry*:

When we reason *a priori*, and consider merely any object or cause, as it appears to the mind, independent of all observation, it never could suggest to us the notion of any distinct object, such as its effect ...

The mistake in this argument is that there is no such thing as the one and only way that an object or cause can appear to the mind. There are many *descriptions* of objects and events, and each of them may be considered “in itself” — in terms of what it logically entails, independent of any auxiliary empirical information. Perhaps the concept of ingesting opium does not, in itself, entail anything about what the effect of ingesting opium would be. But that does not mean that the concept of ingesting something with a

dormitive virtue entails nothing about what the effect of ingesting something with that property would be.³

In summary, there are three types of causal claim — “ e_1 caused e_2 ,” “ e_1 actually promoted e_2 ,” and “ e_1 would promote e_2 .” When “ e_1 occurred” and “ e_2 occurred” are empirical, the first two causal claims are never *a priori* true;⁴ the third sometimes is. Molière’s example is my reason for saying that there are exceptions to Hume’s thesis, but the triviality of this example may suggest that *a priori* statements of the third kind are a mere curiosity. Surely there is no scientific interest in the fact that statements of the form “ e_1 would promote e_2 ” are sometimes *a priori*. But there is — they are at the heart of mathematical theorizing about evolution. Molière’s joke is a joke, but the *a priori* causal models that are part of the theory of natural selection are no laughing matter.

2. A Simple Model

Fitness, like dormitivity and solubility, is a dispositional property, but there is an important difference. The process that leads water-soluble substances to dissolve is usually thought of deterministically⁵ whereas fitness has to do with probabilities.

³ My point about Hume’s M/E argument is consistent with Lewis’ (1986) Humean thesis that one event causes another only if they are “wholly distinct.” Evaluating Lewis’ thesis would require a principle for individuating events, and I have nothing to say on this. But to see what is wrong with Hume’s argument, Davidson’s point is enough: even if the cause of event e is wholly distinct from event e , the fact remains that “the cause of e occurred” *a priori* entails that e occurred.

⁴ Perhaps sentences of the first two types are sometimes *a priori* false. For example, if it is *a priori* true that causes must precede their effects, then “the later event e_1 caused the earlier event e_2 ” is *a priori* false. Of course, it is debatable whether it is *a priori* (or even true) that cause must precede effect. But perhaps there are other *a priori* constraints. For example, if it is *a priori* true that no token event causes itself, then statements of the form “token event e_1 caused token event e_1 ” are *a priori* false.

⁵ If determinism were false, would this mean that nothing is soluble? Perhaps the concept should be understood so that determinism isn’t required for solubility (Sober 1984a, p. 76).

Despite this difference, fitness and solubility are both defined with an eye to the future.

For example, the usual definition of water-solubility goes like this:

x is water-soluble at time $t =_{\text{def}}$ if x were immersed in water (in the right way) at time t , then x would dissolve at some later time t' .

Something similar is true of fitness, as we now will see.⁶

In this section I'll briefly describe a simple model of evolution by natural selection. It has two parts; the model describes the *consequences* of fitness differences and it also describes the *sources* of those fitness differences (Sober 1984a). The consequences of interest are changes in trait frequency; the sources are properties of the relationship of organisms to their environments. These two parts of the model are separable; more sophisticated models often describe one without the other. After sketching this simple model, I'll consider more sophisticated models in the next two sections. All are *a priori*.

Consider a generation of organisms in a population that all begin life at the same time. Every zygote in this generation has trait A or trait B , so the trait frequencies, p and q , sum to 100%. Suppose that individuals with trait A have a probability (w_A) of

⁶ Backwards looking concepts also provide an opening for *a priori* causal claims. Davidson (1987, p. 451) says that "sunburn is caused by exposure to the sun" is *a priori*. If you think he's wrong because tanning machines cause sunburns, let's simply stipulate that the term "schmunburn" is defined as a burn that is caused by exposure to the sun. Now there is no doubting that we have a causal statement that is *a priori*. This example also goes against the Humean grain. Evolutionary biology has its own backwards looking concepts (for example, *homology* and *adaptation*), which also give rise to *a priori* causal statements. Consider, for example, the claim that adaptations for flying are caused by selection for flying (Sober 1984a, p. 208).

surviving to reproductive age while individuals with trait B have a probability (w_B) of surviving. These two probabilities, w_A and w_B , are fitnesses that pertain to viability; they say nothing about how reproductively successful A and B individuals would be if they managed to reach reproductive age. We assume that A and B individuals that manage to develop from egg to adult take the same amount of time to do so.

If $w_A \neq w_B$, natural selection will occur. How will this difference in fitness affect the frequencies of the two traits as the individuals in this single generation develop from egg to adult? The frequencies of the two traits at the adult stage, p' and q' , will have the following expected values:

$$E(p') = pw_A/w \qquad E(q') = qw_B/w$$

Here $w = pw_A + qw_B$; w is the average fitness of the organisms in the population at the egg stage. Notice that it follows from the above two equations that

$$E(p') > p \text{ if and only if } w_A > w_B.$$

In expectation, trait A will increase in frequency precisely when trait A is fitter than trait B .⁷

⁷ The word “expectation” here denotes the mathematical concept by that name, not the subjective state of expecting something to happen. The expected value of a variable is (roughly speaking) the average value it would have over infinitely many trials. If you have a probability of 0.2 of having zero offspring, a probability of 0.7 having exactly one, a probability of 0.1 of having exactly two, and a probability of zero of having more than two, then your expected number of offspring is 0.9, even though you should not expect to have exactly 0.9 babies.

So far the model describes what will happen within a single generation. But the process of natural selection is usually understood as something that occurs over many generations. The model applies to this multi-generational process in a simple way. We imagine that the adults at the end of the generation just described reproduce asexually, with offspring having the traits of their parents. We further imagine that A adults and B adults are equally successful reproductively. There is no fertility selection, only viability selection, in this model. This means that if p' and q' are the frequencies of the two traits at the end of the first generation, then p' and q' are also the frequencies of the two traits at the start of the second. The same process is assumed to operate within this next generation, with the same viability fitnesses applying to the passage from egg to adult. So, at the end of the second generation, the frequencies of the two traits can be expected to change again, and trait A once again will be expected to increase in frequency if A is fitter than B . If the two fitnesses w_A and w_B are constants, trait A can be expected to increase in frequency in each successive generation and eventually to reach 100% representation in the population.

This model can be described without using the causal expression “would promote:”

If A is fitter than B in a population in which no other evolutionary causes are at work, and the traits are perfectly heritable, then A will, in expectation, increase in frequency.

This conditional contains an idealization – the idea that no other evolutionary causes impinge. This idealization can be dispensed with if we simply say: *A's being fitter than*

B would promote A's increasing in frequency if the trait has positive heritability. This formulation should be understood in parallel with the statement that smoking would promote lung cancer. This statement is elliptical; the intended assertion is that smoking would promote lung cancer in individuals of kind *K* (Eells 1991). The US Surgeon General was talking about human beings, not about metal smoking machines or about Martians who thrive on cigarettes. Human beings have a variety of other traits that affect their risk of lung cancer; for example, they may or may not also inhale asbestos. Conjunctions of these other properties comprise “background contexts,” and so the causal claim is elliptical in a second respect. The claim that smoking would promote lung cancer should be understood to mean that smoking would promote lung cancer in individuals of kind *K* relative to a set of background contexts *B*. The intended set *B* rarely includes all possible causal configurations. For example, suppose there is a possible but nonactual genetic mutation that, if it occurred, would lead smoking to reduce the risk of lung cancer. The background contexts contemplated by the Surgeon General probably did not cover this possibility. Once a relevant set of background contexts is circumscribed, we might follow Skyrms (1980) and say that smoking is a positive causal factor for lung cancer if it raises the probability of cancer in some background contexts and does not lower it in any of the others, that smoking is a negative causal factor if it lowers the probability in some but does not raise it in any of the others, and that smoking is a mixed causal factor if it raises in some and lowers in others. This suggestion should be understood as applying to a dichotomous variable (you either smoke or you do not); when there are more than two possible states of a variable, further issues arise (Hitchcock 1993). I mention these definitions here to help fix ideas; my argument does not depend on adopting them.⁸

⁸ Woodward's (2003) manipulationist theory of causation would serve my purposes also. He

The situation is somewhat simpler when we seek to gauge the causal impact that fitness inequalities would have on trait frequencies in a population. If trait *A* were fitter than trait *B*, this would raise the probability of *A*'s increasing in frequency (or raise the expected magnitude of that change) in some relevant background contexts, but not in others. This is what the breeder's equation (Falconer and McKay 1996) says:

response to selection = heritability x strength of selection.

With positive heritability, *A*'s being fitter than *B* would increase the probability of *A*'s increasing in frequency. With zero heritability, *A*'s being fitter than *B* would make no difference. And if the traits have negative heritability,⁹ *A*'s being fitter than *B* would lower the probability of *A*'s increasing in frequency.

I want to understand “*A*'s being fitter than *B*” as a possible cause of *A*'s increasing in frequency. As just noted, other conditions must be satisfied for the increase to occur, but that does not refute the causal claim. If you are squeamish about accepting this causal claim because you think there is a “logical connection” between *A*'s being fitter than *B* and *A*'s increasing in frequency,¹⁰ consider, dear Humean, the following

talks about variables, not the states of variables; a variable *X* is causally relevant to a variable *Y* precisely when there exists a possible manipulation of *X* (the manipulation would change one value of *X* to another) that would be associated with a change in the value of *Y* (or with a change in the probability of values of *Y*). Within this framework, it is possible to talk about a manipulation having positive or negative effect.

⁹ Haldane (1996) recognizes that the mathematics of heritability allows for the possibility of *negative* heritability and cites what he thinks is an empirical example. This idea has never had much currency in biology. I am grateful to Peter Godfrey-Smith for telling me about Haldane's article.

¹⁰ Davidson (1967) addressed a parallel doubt in the philosophy of mind: how can beliefs and desires be causes of action if believing this and desiring that and being rational entail how you

artificial selection experiment. You have several fruitfly populations in different cages. In each, you select for longer wings; in each generation you take individuals with the longest wings and allow them, but not the others, to be the parents of the next generation. After you do this for several generations, you observe that the average wing length in each population has increased. Then you decide to reverse the selection regime. For the next several generations, you choose the flies with the shortest wings to be the parents. You observe that average wing length declines in each population. You have just done a manipulation experiment; you have manipulated whether long is fitter than short. You discover that your manipulations of this variable are associated with whether long increases in frequency or declines. This discovery provides an excellent reason to think that fitness differences cause changes in trait frequencies.¹¹ Of course, the statement that “A’s being fitter than *B* *actually caused* A to increase in frequency” is empirical, because “A is fitter than *B*” is empirical and so is “A increased in frequency.” But “A’s being fitter than *B* *would promote* A’s increasing in frequency” is another matter. It is elliptical for “A’s being fitter than *B* *would promote* A’s increasing in frequency in circumstance *C*.” Circumstances can be specified that make the claim *a priori* true.

Hausman and Woodward (1999, p. 551) assert that variables cannot be causally connected if they are “logically or analytically connected.” They illustrate their point by considering an example in which there are three variables; *X* and *Y* are statistically independent of each other, while a third variable, *Z*, is defined as $Z = X + Y$. Hausman and Woodward say that *Z* will be statistically dependent on *X* and on *Y*, but they deny that *X* and *Y* cause *Z*. I agree. Their conclusion about this example is reinforced if these

will behave? He answered that causal statements can be *a priori* and cited the example discussed earlier to argue for this.

¹¹ Reisman and Forber (2005) describe a manipulation experiment to argue that drift can be a cause of change in trait frequency.

variables describe the state of a system at a single time. If cause must precede effect, X_t and Y_t do not cause Z_t (nor does Z_t cause X_t or Y_t), whether or not the three variables are definitionally connected. In support of their assessment, Hausman and Woodward point out that there is “no conceivable way to intervene on the value of Z ... that is not simultaneously an intervention to change the value of X or to change the value of Y , and there is no way to intervene to change the value of X or to change the value of Y that is not simultaneously an intervention to change the value of ... Z .” This point is specific to the example they are considering. Trait fitnesses at time t_1 and trait frequencies at time t_2 are *not* simultaneous, and it *is* possible to intervene on fitness relations without thereby changing trait frequencies (and *vice versa*). As noted above in connection with the breeder’s equation, if the heritability is zero, changing the fitnesses at time t_1 won’t, in expectation, be associated with any change in the trait frequencies at t_2 . And even when there is positive heritability, intervening on trait frequencies at t_2 won’t be associated with changes in the trait fitnesses at t_1 .¹²

The model described thus far concerns the consequences of fitness differences but not their sources. To start fleshing out the latter idea, I’ll supply a trivial example. Suppose A and B are the traits of running fast and slow in a population of zebras, where predation by lions induces a selection pressure on zebras that favors running fast. We can express this as a conditional:

¹² I don’t have an account to offer of when variables can be related as cause to effect, but that isn’t needed to see that Hausman and Woodward’s point about their X , Y , and Z doesn’t generalize to the variables involved in models of natural selection.

If lions were to hunt and kill slow zebras more successfully than they hunt and kill fast ones, and if this were the only source of fitness differences in the zebra population, then fast zebras would be fitter than slow ones.

Here again, the idealization in the antecedent can be removed by using causal language: *if lions were to hunt and kill slow zebras more successfully than they hunt and kill fast ones, then running fast would promote survival.* This does not mean that fast zebras *will* be fitter; after all, fast zebras may also have longer and thinner leg bones, and these might be more prone to breaking (Maynard Smith 1978).

The simple model just outlined describes a source and a consequence of fitness differences. The model can be expressed by conditionals that use idealizations, but the idealizations can be dropped by using the locution “would promote.” The model is *a priori*, whichever way it is represented. Of course, the existence of zebras and lions isn’t *a priori*, nor is the fact that the population of zebras persists through many generations, nor is the fact that the traits in question are heritable. But the model does not entail any of these. Like the empirical laws found in other sciences, the present model simply says *what would happen if*. The model is dynamical, meaning that it describes how a system would change through time if it were subject to an array of causal influences. Humeans maintain that dynamical causal models are always empirical. The model just described is an exception.

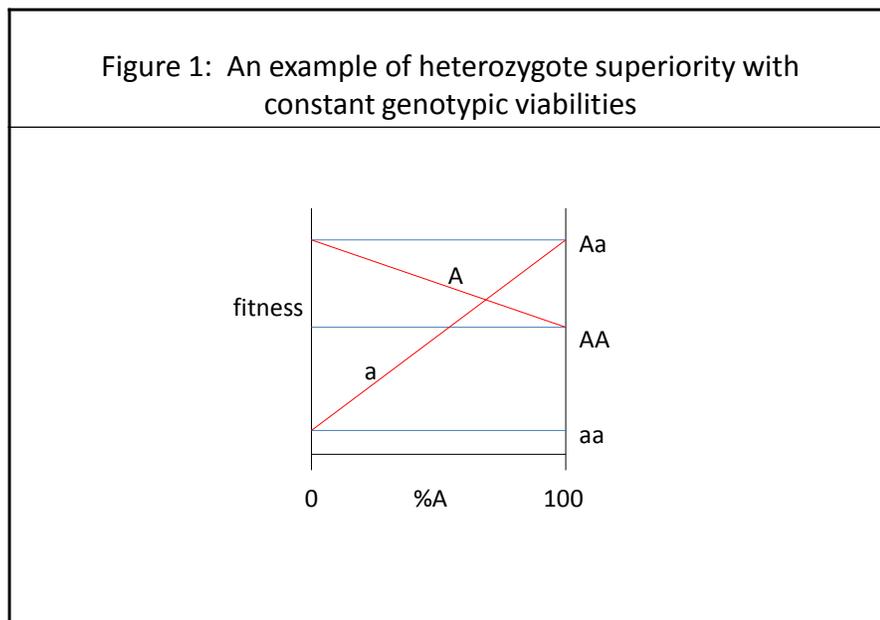
3. A More Sophisticated Consequence Model

Even if the simple model just described is *a priori*, perhaps matters must change once more sophisticated models are examined. This is not so, as I will try to demonstrate in this section and the next one.

In the constant viability model described in §2, selection is a destroyer of variation. The population begins with a mix of *A* and *B* individuals, and selection then drives the fitter trait to 100%. In contrast, there is a simple genetic model in which selection preserves variation rather than destroying it. It is from Evolution 101, but should interest those who think that the consequence model described earlier is all there is to this part of mathematical biology.

Suppose there are two alleles (*A* and *a*) that can exist at a locus in the diploid organisms in a population, so there are three possible genotypes (*AA*, *Aa*, and *aa*). Let the fitnesses of these three genotypes be w_1 , w_2 , and w_3 , respectively, and suppose that heterozygotes (the *Aa* individuals) are fitter than homozygotes. As before, let's consider a single generation that begins with the two genes, *A* and *a*, at frequencies p and q respectively; the genotypes *AA*, *Aa*, and *aa* have frequencies p^2 , $2pq$, and q^2 . The frequencies of the three genotypes after selection are, in expectation, $p^2 w_1/w$, $2pq w_2/w$, and $q^2 w_3/w$, where the normalizing denominator w is again the average fitness of individuals in the population; $w = p^2 w_1 + 2pq w_2 + q^2 w_3$. As was true earlier, the fittest of the traits will increase in frequency in the passage from egg to adult, so the heterozygote form will increase in frequency within this generation. If this is true in *every* generation and there is no fertility selection, will heterozygotes go to 100% in the population? The answer is *no*; Mendelian reproduction prevents this from happening.

When heterozygotes mate with each other, their offspring can be expected to exhibit all three genotypes. What happens in this model is that the gene frequencies evolve to an equilibrium in which both alleles and all three genotypes are represented. This equilibrium is deducible just from the three genotypic fitnesses; at equilibrium, the ratio of p to q takes the value $(w_2 - w_3)/(w_2 - w_1)$. Figure 1 provides an example that illustrates the qualitative properties of this model. The fitnesses of the three genotypes are frequency independent. You can deduce from them what the fitnesses of the two alleles are. The allelic fitnesses are frequency dependent; they have the property that a rare allele is fitter than a common one. The equilibrium is the allelic frequency at which the two alleles have the same fitness.



As before, we can describe this model by writing a conditional that contains an idealization and we can remove the idealization by using the locution “would promote.”

The idealization reads: if there were heterozygote superiority in a large population and random mating and no other evolutionary causes impinged, then a stable polymorphism would probably evolve. The causal locution allows us to be more concise: heterozygote superiority would promote the evolution of a stable polymorphism. The model is *a priori*, though of course determining whether it applies to this or that locus in this or that population is an empirical task.

5. A More Sophisticated Source Model

It is pretty obvious that running fast would promote zebra survival if zebras were hunted by lions. Those unfamiliar with evolutionary biology may think that theorizing about the sources of fitness differences is always a no-brainer of this sort. It isn't, as my next example will show.

Why are there approximately equal numbers of males and females in human populations? And what is the explanation of the sex ratios found in the rest of nature? Darwin advanced a theory to explain this in the first edition of the *Descent of Man*. The idea was that if a population is monogamous and has an uneven sex ratio, selection will favor the overproduction of the minority sex (Darwin 1871, p. 316). For example, if there are more males than females, then the best reproductive strategy for a parent is to produce all daughters (since all of them will find mates). Symmetrically, if there are more females than males, the best strategy is to have all sons. The result is that a population that has an uneven sex ratio will evolve towards 50/50. Darwin notes that the logic of this argument also applies to polygyny. If the mating groups in a population consist of 1 male and 3 females, and there are more than 75% females in the population,

then selection will reduce the female bias to 75%, at which point selection will cease (Darwin 1871, p. 317).

This may seem like a pretty good theory, but Darwin withdrew it in the book's second edition (Darwin 1874, pp. 267-268). He does not explain why.¹³ The reason he was right to do so became clear some years later when Carl Düsing advanced a very different theory. Düsing (1884) was able to show that monogamy and polygyny do not matter. Rather, the key is that each offspring has one mother and one father. Düsing's model requires us to consider three generations – (1) parents, (2) offspring, and (3) grandoffspring. Suppose there are m males and f females in the offspring generation, and N individuals in the generation of grandoffspring. This means that the average male in generation 2 has N/m offspring in generation 3 and the average female has N/f . If a parent in generation 1 is to maximize the number of grandoffspring she has in generation 3, what mix of sons and daughters should she produce in generation 2? The answer is that she should produce offspring who are solely of the minority sex. This is true regardless of whether there is monogamy or polygyny. If there is polygyny, some males are very successful in reproducing while others are not. Düsing's insight was to focus on the *average* male and the *average* female, not just on the unlucky individuals who fail to mate. Darwin reasoned incorrectly about this problem.

But Düsing got a detail wrong. He, like Darwin, thought his theory predicted the sex ratio that should exist at reproductive age. R.A. Fisher (1930) saw that the argument in fact predicts what the sex ratio should be at the age of independence; this is where Düsing's N/m and N/f really apply. Besides correcting this mistake in Düsing's

¹³ Sober (2010, §3.4) speculates that Darwin withdrew his theory because he suspected that even sex ratios evolved without monogamy's being in place.

reasoning, Fisher placed Düsing's argument in a more general setting. Fisher did so by introducing a new concept — the idea of *parental expenditure*. A parent has a package of energy to invest in the creation and rearing of offspring. She divides that package, devoting some of it to daughters and the rest to sons. An offspring may die before the age of independence, and after the age of independence he or she may die before reaching reproductive age. Fisher's main result was that the sex ratio in a population should evolve to the point at which there is *equal investment* in sons and daughters. If the average son costs c_s units of energy to create and rear and the average daughter costs c_d , then what should evolve is the equality $mc_s = fc_d$. If the two sexes are equally costly, an even sex ratio will evolve. If sons are less costly, a male-biased sex ratio should evolve. The latter is what we observe in human populations; sons are less costly than daughters because they have higher mortality rates before the age of independence.

There is an assumption in Fisher's model, and in Düsing's also, that was pretty much invisible until Hamilton (1967). This is the assumption of random mating; the males and females in generation 2 mate at random to produce generation 3. Hamilton considered a situation in which mating is nonrandom. He imagined a hypothetical species of parasitic wasp in which a single fertilized female lays her eggs inside a host. The eggs hatch and the siblings that emerge then mate with each other, after which the fertilized females disperse to find new hosts. There is inbreeding here, not random mating. What will evolve in this circumstance is a female-biased sex ratio. Hamilton showed that the smaller the number of foundresses parasitizing a single host, the more female-biased the sex ratio will be. As the number of foundresses per host increases,

Hamilton's prediction moves towards Fisher's. Just as Fisher generalized Düsing, so Hamilton generalized Fisher.¹⁴

I have loosely described these different models as predicting what sex ratios should evolve. They do make this type of prediction (assigning the predicted frequencies a probability = 1) if the population is infinite and the traits in question have positive heritability and no other evolutionary causes impinge. But we can dispense with these idealizations and focus just on what these models say about the sources of fitness differences. Let's consider a single consequence of Fisher's model. Suppose there are two reproductive strategies that a parent can follow; she can produce all sons (*ALLS*) or she can produce all daughters (*ALLD*). Fisher's model says, of a random mating population, that $mc_s > fc_d$ would promote the evolution of *ALLD*. Once spelled out carefully, this generalization is *a priori*.

7. Objection #1: Trivialization

I am not claiming that the pattern I have described is unique to natural selection. It also applies to other evolutionary processes; for example, the consequence models that describe the process of random genetic drift are *a priori* as well. And perhaps there is *a priori* causal modeling in other sciences; maybe economics furnishes examples, with rational agents maximizing utility being an approximate analog of organisms differing in fitness. What I do claim is that the pattern to which I have been pointing is far from universal. For the most part, scientific statements about causal relations are empirical; Hume was right most of the time.

¹⁴ A relation of *reduction* is instantiated: Düsing's model reduces to (is a special case of) Fisher's, and Fisher's bears that relation to Hamilton's.

The objection I want to consider asserts that the pattern I have described in evolutionary theory applies trivially to all scientific theories. For example, consider Newtonian theory (mechanics + gravitation), which is generally taken to be empirical. Since this theory (*T*) allows one to deduce a prediction (*P*) from a specification of initial and boundary conditions (*IB*), we can construct a conditional of the form “If *IB* and *T*, then *P*” that is *a priori* true. The objection is correct that *T*’s being empirical does not prevent this construction from being carried out. But the situation in evolutionary theory is different. The *a priori* models I have described have the form “if *IB*, then *P*”; there is no empirical law in the antecedent. For example, the antecedent of the conditional I described in connection with Fisher’s model of sex ratio evolution in §3 says that the population is random mating, that various sex ratio strategies are represented, that there is over-investment in sons, and that the traits have positive heritability. None of these are empirical laws. True, they are empirical, but they describe contingent features exhibited by some populations but not by others. The same point holds for the antecedent of the conditional that expresses the model of heterozygote superiority described in §3. This antecedent also includes the claim that there is random mating. In addition, it says that the genes in the population obey Mendel’s (so-called) laws. Beatty (1995) is correct that Mendel’s “laws,” when true, are contingent outcomes of the evolutionary process.

A related objection to my thesis can be formulated by using Giere’s (1979) distinction between what I’ll call “model specifications” and “model applications.” The former are definitions; for example, we can define a Mendelian genetic system as any collection of objects that obeys a given set of generalizations and a Newtonian particle system can be defined in the same way. These definitions are *a priori*. The second part of Giere’s picture of science involves hypotheses that say that this or that real world

system is a Mendelian genetic system or a Newtonian particle system. Within this framework, there is nothing in science that is both general and empirical; there are general *a priori* claims and singular empirical claims, and that is it. This bifurcation may have its uses, but it does not constitute an objection to what I have claimed. It misses the point that there are theories (like Newton's) that are both general and empirical and also the fact that such theories differ from the models of natural selection that I have described.

8. Objection #2: Where's the Causality?

In discussing my three examples of models of natural selection, I gave each of them a pair of formulations. The first was in terms of a conditional; the second used the locution "would promote." An objector might grant that the first of these is *a priori*, but deny that the second describes a consequence of the model or express puzzlement about what this second formulation means or why we should think it is true. My reply is to return to the manipulation experiment described in §2. The model of heterozygote superiority (to focus on this one example) tells you how to compute the gene frequencies that would, in expectation, evolve, given a set of genotypic fitnesses and a set of initial and boundary conditions. But it also describes how manipulating genotypic fitnesses in populations of a certain kind would tend to be associated with different evolutionary outcomes. This is what justifies describing these models by using the causal locution of "would promote."

9. Objection #3: "x would promote y" isn't ever *a priori* because ...

Alan Hájek formulated the following objection to my claim that "x would promote y" is sometimes *a priori*. The objection is that the statement entails that y is

promotable, and it is never *a priori* that *y* is promotable.¹⁵ My reply is to deny the entailment. The statement “*x* would promote *y*” could be true even if it is impossible to promote *y*; of course, if *x* would promote *y* and *y* can’t be promoted, then it follows that *x* isn’t possible. Running a 3 minute mile would promote your becoming world famous. But that doesn’t mean that it is possible for you to become world famous, or that you could ever run that fast.

10. What’s Empirical in the Theory of Natural Selection?

The causal generalizations that I say are *a priori* have the form “*x* would promote *y*.” The two other types of causal claim that I discussed – “*x* actually promoted *y*” and “*x* actually caused *y*” – are not. These last two also are important in evolutionary theory and it is in deploying them that the theory goes empirical.

A nice example of this empirical turn is furnished by Darwin’s reasoning in the *Origin* about why young mammals (including human babies) have skull sutures. Darwin recognizes that live birth *would promote* the evolution of skull sutures. His question is whether live birth *actually promoted* the evolution of this trait. Darwin’s answer is *no*:

The sutures in the skulls of young mammals have been advanced as a beautiful adaptation for aiding parturition, and no doubt they facilitate, or may be indispensable for this act; but as sutures occur in the skulls of young birds and reptiles, which have only to escape from a broken egg, we may infer that this

¹⁵ I won’t dispute promotability’s always being an empirical matter. The modality in “can be promoted” isn’t logical possibility; usually, it’s some sort of nomological possibility in the circumstances.

structure has arisen from the laws of growth, and has been taken advantage of in the parturition of the higher animals (Darwin 1859, p. 197).

Given that Darwin's question is why *mammals* now have skull sutures, why does it matter that *nonmammals* have skull sutures but not live birth? Darwin is not being sidetracked by an irrelevancy. He is reasoning phylogenetically. He is saying that the fact that reptiles and birds have sutures but not live birth is evidence that sutures predated live birth in the lineage leading to modern mammals (Sober 2010, §5.2). This is why Darwin thinks it is false that live birth *actually* promoted the evolution of skull sutures in this lineage.

Darwin mentions “the laws of growth” in explaining why skull sutures exist, but his argument can be understood without invoking that idea. We can take Darwin to be testing two hypotheses concerning what actually promoted what. The observation (*O*) that birds and reptiles have skull sutures but no live birth discriminates between the two hypotheses:

$\Pr(O \mid \text{live birth actually promoted the evolution of skull sutures in the lineage leading to modern mammals}) <$

$\Pr(O \mid \text{skull sutures actually promoted the evolution of live birth in the lineage leading to modern mammals}).$

This representation of Darwin's argument makes use of the Law of Likelihood (Hacking 1965, Sober 2008). There is more to reasoning about natural selection than recognizing that one trait *would have promoted* the evolution of another. There is the further issue of *actual promotion*.

Just as “ x would promote y ” does not entail that x actually promoted y , so “ x actually promoted y ” does not entail that x actually caused y . It isn’t just that “ x actually promoted y ” does not entail that y occurred, but “ x actually caused y ” does. There is a further gap. Suppose Smith smoked cigarettes and inhaled asbestos and that both actually promoted his getting lung cancer. Suppose he then gets lung cancer. The question remains of whether his cancer was actually caused just by his smoking, or just by his inhaling asbestos, or by both, or by neither (Sober 1984b; Woodward 1990, 1994; Hitchcock 2004).¹⁶

If x and y both occurred and x actually promoted y , what more is needed for x to have caused y ? I don’t have an answer to this metaphysical question,¹⁷ but something can be said about its epistemology. Perhaps when smoking causes cancer, the process of carcinogenesis is different from the process that takes place when asbestos causes lung cancer (Hitchcock 2004; see also Sober 1984a and 1999). For example, suppose that when smoking causes lung cancer, the smoking causes an S mutation, which then causes the tumor. And suppose that when asbestos causes lung cancer, the asbestos causes an A mutation, which then causes the tumor. Another possibility is that when smoking causes lung cancer, the process has side effects that differ from those that occur when asbestos causes lung cancer. We can gain evidence about these sorts of facts by observing

¹⁶ Not only does “ x promoted y ” not entail that x actually caused y . The converse entailment fails as well. The reason is that “ x promoted y ” requires x to raise the probability of y ; but x can be a cause of y even though x didn’t raise y ’s probability; y still can “trace back” to x (Sober 1984b).

¹⁷ Woodward (2003, pp. 74-86) offers an account of actual causation that assumes that causes necessitate their effects; he does not address the question of what actual causation means in an indeterministic setting.

carcinogenesis in individuals who have smoked but have not been exposed to asbestos, and in individuals who have been exposed to asbestos but have not smoked. The more general point is that an observation O will favor one causal hypothesis over another if it induces an inequality of likelihoods:

$$\Pr(O \mid \text{Jones' smoking, but not his asbestos exposure, caused his lung cancer}) >$$

$$\Pr(O \mid \text{Jones' asbestos exposure, but not his smoking, caused his lung cancer}).$$

Here again the Law of Likelihood comes into play. The observation that Jones got lung cancer does not discriminate between the two causal hypotheses, since it is entailed by each. But other observations may do so.

Given the distinction between “ x promoted y ” and “ x actually caused y ,” do evolutionary biologists bring evidence to bear, not just on the former, but on the latter? They do, and Darwin’s discussion of skull sutures provides an example of this as well. The observation O , that reptiles and birds have skull sutures but no live birth, discriminates between two hypotheses about actual causation:

$$\Pr(O \mid \text{in the lineage leading to modern mammals, skull sutures caused live birth to evolve}) >$$

$$\Pr(O \mid \text{in the lineage leading to modern mammals, live birth caused skull sutures to evolve}).$$

The traits of modern birds and reptiles are “side effects” of what happened in the lineage leading to modern mammals. These side effects are evidentially relevant.

Although this example shows that biologists do test hypotheses about actual causation against each other, the example is not one in which they address a question that is structurally the same as the one about Jones' lung cancer. What is wanted is a case in which there is an evolutionary outcome that was actually *promoted* by two or more token events, but was actually *caused* by only some of them. There are many such examples in evolutionary biology. I'll briefly describe three.

Suppose, as before, that there are two alleles and three genotype at a locus, but now let's consider what happens when heterozygotes are exactly intermediate in fitness, with $w_{AA} > w_{Aa} > w_{aa}$. If the population begins with the *a* allele at 100% and a single mutation from *a* to *A* occurs, what is the probability that this new allele will sweep to 100%? The answer is that the probability of fixation goes up as $N_e s$ increases, where s is the fitness difference between *Aa* and *aa* and N_e is the effective population size. The favorable mutant will probably be lost before it reaches fixation if $N_e s$ is small; for example, if $N_e s = 1$, the probability of fixation is about 0.004 (Hamilton 2009, p. 240). In such cases, evolving from 100% *a* to 100% *A* becomes probable when there are *repeated* mutations from *a* to *A*. Each mutation will probably fail to reach fixation, but the probability is high that at least one will succeed. Suppose *A* does evolve to 100% in the population we are considering. We can look back on the process and say the following: Each mutation from *a* to *A* *promoted* the evolution of 100% *A* (because each mutation raised the probability of that outcome). But the evolution of 100% *A* probably *traces back* to a small fraction of the favorable mutations that occurred; they, and not the other mutations of the same type, are the *actual causes* of the trait's reaching 100%.

In this example, there are several token events that promoted the evolution of 100% *A* in the population, but they are all of the same type – all are mutations from *a* to *A*. For this

reason, the present example about natural selection is a bit disanalogous with the one about smoking and asbestos both promoting Jones' lung cancer. In the Jones example, the two promoters differ in kind. This imperfect parallel is easily corrected. Suppose there is selection for longer wings in a population of fruitflies and that the population evolves from 100% short wings to 100% long. The mutations that promoted this change can be genetically different, not only in terms of their nucleotide sequences, but in terms of the developmental pathways they help set in motion. Suppose the population initially has short wings because all individuals are homozygous for the allele a_0 and that average wing length would go up if this allele were replaced by a_1 or by a_2 or ... by a_n or by any combination of these. Each mutation from a_0 to a_i ($i > 0$) promotes the evolution of 100% long wings. If the process of evolving from short wings to long spans a large number of generations and the population is large, many such mutations will occur. After long wings reach fixation in the population, suppose we discover that this phenotype is now universal because all individuals are homozygous for a_1 . Suppose the long wings at this later time trace back to one or more mutations from a_0 to a_1 , but not to the mutations that occurred from a_0 to a_i ($i > 1$). These other mutations actually promoted the evolution of 100% long wings, but they did not cause it.

This example about wing length makes an epistemological question easier to answer that may earlier have seemed intractable. In the first example I described, repeated mutations from a to A each promoted the evolution of 100% A , though it is probable that only a few of these mutations actually caused that outcome. The question then arises of whether we can know which of these mutations were actual causes. Maybe this is difficult. But in the new setting in which a phenotypic change can be caused by different kinds of genetic change, the question is easier. If the population evolved from 100% short wings to 100% long, what

genetic changes actually caused this change? Examining the genes present at the start and at the finish can help answer that question.¹⁸

A third biological example that illustrates the distinction between “ x actually promoted y ” and “ x actually caused y ” comes from Mendelian inheritance. Consider an Aa heterozygote in the present generation and its two heterozygote parents. Figure 2 describes how the offspring’s being a heterozygote depends on the genotypes of both parents. The table is symmetrical; mom and dad are identical in the way they probabilify the offspring’s genotype. There is an oddity in this example: if one parent is a heterozygote, the genotype of the other parent does not affect the probability that the offspring will be a heterozygote. This is a probabilistic analog of overdetermination. If both parents are heterozygotes, it takes a double manipulation of the parental genotypes to make a difference in the offspring’s probability of being Aa ; changing just one of them won’t have any effect.

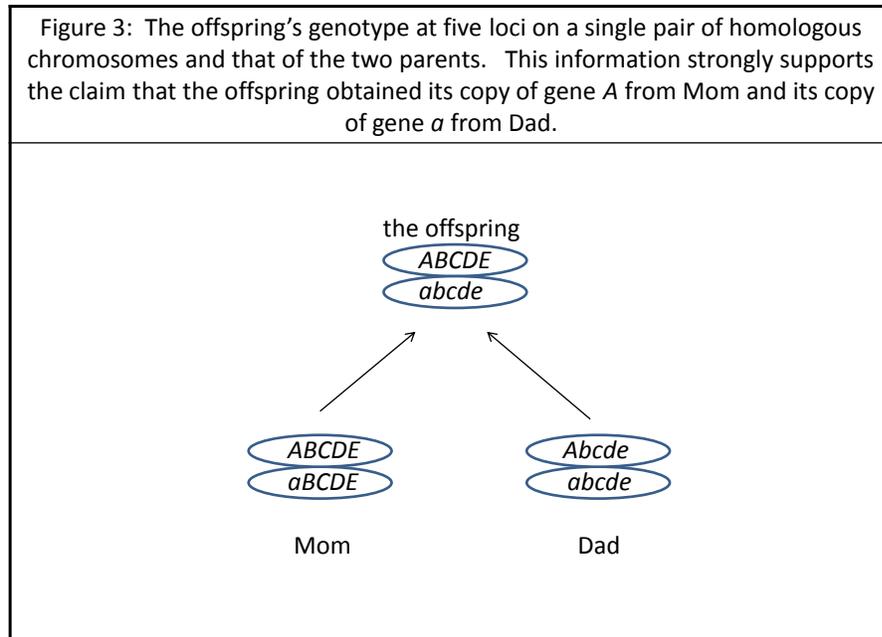
¹⁸ Absent a time machine, you can’t now observe the genetic composition that a present population had in the past; however, you can examine other contemporary populations that share common ancestors with that present population. These collateral descendants (Darwin’s term) provide evidence about the genetic characteristics of ancestors (see Sober 2010, §1.6 and §5.1, for discussion).

Figure 2: The probability of the offspring's being a heterozygote depends on the genotypes of the two parents.

| | | Dad is | | |
|--------|----|---------------|---------------|---------------|
| | | AA | Aa | aa |
| Mom is | AA | 0 | $\frac{1}{2}$ | 1 |
| | Aa | $\frac{1}{2}$ | $\frac{1}{2}$ | $\frac{1}{2}$ |
| | aa | 1 | $\frac{1}{2}$ | 0 |

In spite of this symmetry between Mom's and Dad's contributions to the offspring's being Aa , there is an asymmetry. One parent *contributed* allele A and the other contributed allele a ; the A allele in the offspring *traces back* to one parent, not the other. How might this fact about the actual construction of the offspring genotype be inferred? Although the mere fact that the offspring is Aa does not provide any evidence about which allele came from which parent if both parents are Aa as well, other facts about the offspring and the two parents might be more telling. For example, consider other loci that are adjacent to the A -locus and the genes that are present there, as shown in Figure 3. Suppose the offspring has genes $BCDE$ next to his copy of A and has genes $bcde$ next to his copy of a ; the offspring is a heterozygote, not just at the A locus, but at these four adjacent loci as well. Suppose further that Mom is homozygotic for $BCDE$ at these other loci, while Dad is homozygotic for $bcde$. This is strong

evidence that the offspring obtained *A* from Mom and *a* from Dad.¹⁹ We now have a pretty good parallel to the case in which we have evidence that Jones' smoking, not to his asbestos exposure, was what actually caused his lung cancer — that evidence being that there are other, correlated, effects that we'd expect to observe if Jones' cancer traced back to his smoking, but would expect not to find if the cancer traced back to asbestos exposure.



11. Concluding Comment

I suspect that philosophers often under-value the role of *a priori* truths in empirical science because they think that *a priori* truths are trivial. Apart from its technical meaning in logic, the term “tautology” is pejorative.²⁰ Never mind the fact that

¹⁹ How could *A* come from Dad and *a* from Mom? One possibility is that eight specific mutations occurred in the gametes from Mom and Dad that produced the offspring. This is not impossible, but it is less probable than that none of these mutations occurred.

²⁰ An example is the accusation that evolutionary theory is a “tautology.” The main point that needs to be made in reply to this criticism is independent of whether causal claims can be *a priori*.

tautologies are *true*; the point is that they are supposed to be “empty.” For those who find the idea of the synthetic *a priori* unattractive, the *a priori* tends to suggest examples like “bachelors are unmarried men;” such statements merely provide definitional abbreviations and furnish zero insight into the nature of reality. The flip side of this picture is the Humean thesis that causal claims and laws of nature, because they are nontrivial, must be empirical.²¹

“Bachelors are unmarried men” is not a causal claim and it would be silly to call it a law of nature. The mistake is to think that this point generalizes — that laws are always empirical and that causal claims always are too. If *a priori* generalizations figure in explanations and predictions in the same way that empirical laws do, we should regard these *a priori* generalizations as laws (Elgin 2003). And some causal claims are *a priori*. Molière’s joke provides a trivial example of an *a priori* causal claim. Mathematical models of natural selection furnish non-trivial examples. Hume was wrong that laws and causal claims are always empirical.²²

University of Wisconsin, Madison

Even if there are various definitional truths in evolutionary biology, there is lots more that is empirical (Sober 1984a).

²¹ In the *Enquiry*, Hume (Section IV, Part 1) expresses his thesis that causal claims are empirical and his thesis that laws of nature must be empirical in nearly the same breath: “all the laws of nature, and all the operations of bodies without exception, are known only by experience.”

²² I am grateful to John Basl, David Braddon-Mitchell, Craig Callendar, Christopher Hitchcock, Mark Colyvan, James Crow, José Diez, Antonio Dieguez, Mehmet Elgin, Patrick Forber, Mathias Frisch, Martha Gibson, Michael Goldsby, Peter Godfrey-Smith, Paul Griffiths, Alan Hájek, Dan Hausman, Carl Hofer, Jack Justus, Leon Leontyev, Sarah Paul, Huw Price, Alex Rosenberg, Alan Sidelle, Lionel Shapiro, Alan Sidelle, Dennis Stampe, Kim Sterelny, Mauricio Suarez, Michael Titelbaum, Ken Waters, and James Woodward for helpful comments. I also am grateful to the Sydney Centre for the Foundations of Science for its hospitality during March and April 2010.

References

- Beatty, J. 1995. "The Evolutionary Contingency Thesis, in *Concepts, Theories, and Rationality in the Biological Sciences*, ed. G. Wolters and J. Lennox, Pittsburgh: University of Pittsburgh Press: 45-81.
- Darwin, C. 1859. *The Origin of Species*. London: Murray; Cambridge: Harvard University Press, 1964.
- Darwin, C. 1871. *The Descent of Man and Selection in Relation to Sex*. London: Murray, first edition.
- Darwin, C. 1874. *The Descent of Man and Selection in Relation to Sex*. London: Murray, second edition.
- Davidson, D. 1963. Actions, Reasons, and Causes, in *Essays on Actions and Events*, Oxford: Oxford University Press, 1980, 3-20.
- Davidson, D. 1987. Knowing One's Own Mind, *Proceedings and Addresses of the American Philosophical Association* 60/3: 441-458.
- Düsing, C. 1884. *Die Regulierung des Geschlechtsverhältnisses*. Jena: Gustav Fischer. Selections translated in Edwards (2000).
- Edwards, A. 2000. Carl Düsing on the Regulation of the Sex-Ratio, *Theoretical Population Biology* 58/3: 255-257.

Eells, E. 1991. *Probabilistic Causality*, Cambridge: Cambridge University Press.

Elgin, M. 2003. Biology and *A Priori* Laws, *Philosophy of Science* 70/5: 1380–1389.

Falconer, D. and MacKay, T. 1996. *Introduction to Quantitative Genetics*. Edinburgh: Pearson.

Fisher, R. 1930. *The Genetical Theory of Natural Selection*. New York: Dover, 2nd edition, 1958.

Giere, R. 1979. *Understanding Scientific Reasoning*. New York: Holt, Rinehart & Winston.

Hacking, I. 1965. *The Logic of Statistical Inference*. Cambridge: Cambridge University Press.

Haldane, J. 1996. The Negative Heritability of Neonatal Jaundice, *Annals of Human Genetics* 60/1: 3-5.

Hamilton, M.B. 2009. *Population Genetics*, London: Wiley-Blackwell.

Hamilton, W. 1967. Extraordinary Sex Ratios, *Science* 156/3774: 477-488.

Hausman, D. and Woodward, J. 1999. Independence, Invariance, and the Causal Markov Condition, *British Journal for the Philosophy of Science* 50/4: 521-583.

Hitchcock, C. 1993. A Generalized Probabilistic Theory of Causal Relevance, *Synthese* 97/3: 335-64.

Hitchcock, C. 2004. Do All and Only Causes Raise the Probabilities of Effects? in *Causation and Counterfactuals*, ed. J. Collins, N. Hall, and L. Paul, Cambridge: MIT Press: 403-418.

Lewis, D. 1986. Events, in *Philosophical Papers*, vol. 2, Oxford: Oxford University Press: 241-270.

Maynard Smith, J. 1978. Optimization Theory in Evolution, *Annual Review of Ecology and Systematics* 9: 31-56.

Reisman, K. and Forber, P. 2005. Manipulation and the Causes of Evolution. *Philosophy of Science* 72/5: 1113–1123.

Skyrms, B. 1980. *Causal Necessity: a Pragmatic Investigation of the Necessity of Laws*. New Haven: Yale University Press

Sober, E. 1984a. *The Nature of Selection*. MIT Press.

Sober, E. 1984b. Two Concepts of Cause. In *PSA 1984, Proceedings of the Biennial*

Meeting of the Philosophy of Science Association, ed. P. Asquith and P. Kitcher,
2: 405-424.

Sober, E., 1999. The Multiple Realizability Argument against Reductionism, *Philosophy of Science* 66: 542-564.

Sober, E. 2008. *Evidence and Evolution – the Logic Behind the Science*. Cambridge University Press.

Sober, E. 2010. *Did Darwin write the Origin Backwards?* Prometheus Books.

Woodward, J. 1990. Supervenience and Singular Causal Statements. In *Explanation and its Limits*, ed. D. Knowles, Cambridge: Cambridge University Press: 215-216.

Woodward, J. 1994. Review of Humphreys *The Chances of Explanation*, *British Journal for the Philosophy of Science* 45/1: 353-374.

Woodward, J. 2003. *Making Things Happen*. Oxford: Oxford University Press.