

# Fitness and the Twins

Elliott Sober\*

---

Michael Scriven's (1959) example of identical twins (who are said to be equal in fitness but unequal in their reproductive success) has been used by many philosophers of biology to discuss how fitness should be defined, how selection should be distinguished from drift, and how the environment in which a selection process occurs should be conceptualized. Here it is argued that evolutionary theory has no commitment, one way or the other, as to whether the twins are equally fit. This is because the theory of natural selection is fundamentally about the fitnesses of traits, not the fitnesses of token individuals. A plausible philosophical thesis about supervenience entails that the twins are equally fit if they live in identical environments, but evolutionary biology is not committed to the thesis that the twins live in identical environments. Evolutionary theory is right to focus on traits, rather than on token individuals, because the fitnesses of token organisms (as opposed to their *actual* survivorship and degree of reproductive success) are almost always unknowable. This point has ramifications for the question of how Darwin's theory of evolution and R. A. Fisher's are conceptually different.

---

## Keywords

Darwin, Charles • drift • evolution • Fisher, R. A. • fitness • natural selection • operationalism • supervenience

## 1 Introduction

Michael Scriven's (1959) example of the identical twins who do unequally well in the project of surviving and reproducing has long occupied a central place in the philosophy of biology literature on fitness and natural selection.<sup>1</sup> In Scriven's story (and in the variants of it that other philosophers have used), you are asked to imagine that the twins have identical genotypes

---

<sup>1</sup>See, for example, Abrams (2009a, 2009b), Beatty (1980, 1984), Beatty and Finsen (1989), Brandon (1978, 1990, 2005), Drouet and Merlin (2015), Ettinger et al. (1990), Hodge (1987), Hull (1974), Kimbrough (1980), Matthen and Ariew (2002), Michod (2000), Mills and Beatty (1979), Millstein (2002, 2016, 2017), Pfeifer (2005), Ramsey (2006), Rosenberg (1985), Shanahan (1992), Sober (1984, 2000), Sterelny and Kitcher (1988), Stephens (2004, 2014), and Triviño and Nuño de la Rosa (2016).

---

\*Department of Philosophy, 5185 Helen C. White Hall, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA, ersober@wisc.edu

Received 31 May 2019; Revised 8 September 2019; Accepted 8 September 2019  
doi:10.3998/ptpbio.16039257.0012.001



and phenotypes,<sup>2</sup> and live in the same environment. One of them is childless and is killed by lightning (in Scriven's telling, the death is due to a bomb exploding or a tree falling), while the other isn't hit by the lightning and goes on to have some offspring. The philosophical lesson drawn from this example can be encapsulated in a simple argument:

- (F) The twins are equal in fitness.  
The twins are unequal in their degrees of reproductive success.  
 Hence, fitness is not the same property as actual reproductive success.

A further conclusion is then drawn: fitness is the *ability* an organism has to survive and reproduce. The twins were equally able, but unequally lucky. When the ability they share is characterized probabilistically, it is called a "propensity." The propensity interpretation of fitness is now a standard part of how many philosophers of biology think about evolutionary theory,<sup>3</sup> but see Sober (2013) and Drouet and Merlin (2015) for dissenting verdicts.

The example of the twins has also been used to address the question of how natural selection and drift are related. One influential picture is that drift is "indiscriminate sampling" and selection is "discriminate sampling;" that is, drift occurs when fitnesses are equal and selection occurs when they are not (Beatty 1984; Hodge 1987; Millstein 2002).<sup>4</sup> Lightning in Scriven's example is said to induce indiscriminate sampling, not selection.<sup>5</sup>

In this paper, I'll argue that the first premise of the F argument, which says that the twins are equally fit, is not something that evolutionary theory obliges us to accept. I do agree, though, that the argument's conclusion is correct, so a better argument for it is needed. I try to supply that better argument in what follows.

Before I get started, I want to comment on my choice of terminology. I'll talk about the fitnesses of token individuals and the fitnesses of traits. I won't use the phrase "individual fitness," since that wording encourages the very ambiguity I want to avoid. The term sometimes is used to refer to the fitness of a trait (like running fast) that attaches to individuals, but it sometimes is used to refer to the fitnesses of unique individuals (like Joe). The philosophical terminology of "token" and "type" is handy here; it avoids this ambiguity, which also is pervasive in ordinary language. If someone says that you and I own "the same computer," this might mean that there are two token computers of the same type, and we each own one of them, or it might mean that there is a single token computer, and we each are co-owners of it. Traits are types; they are instantiated in some number (0, 1, 2, ...) of token individuals.

<sup>2</sup>Monozygotic twins often fail to be genetically identical; after a fertilized egg divides, mutations sometimes arise in one twin but not the other (Casselman 2008), and monozygotic twins almost always develop different phenotypes. So the story about the "identical" twins is extremely hypothetical (not that there's anything wrong with that).

<sup>3</sup>It's a further step to say that the propensity in question can be defined in terms of a *mathematical expectation*—the expected number of offspring the individual will have. This proposal faces two problems. First, fitness isn't always defined in terms of next-generation expectations (Beatty and Finsen 1989); for example, the expected number of *grand*offspring matters in sex-ratio theory (Sober 2011). Second, variance in offspring number is known to matter to fitness; the mathematical expectation is not sufficient (Gillespie 1977; Frank and Slatkin 1990; Brandon 1990; Sober 2001). Pence and Ramsey (2013) address both of these topics.

<sup>4</sup>This formulation does not describe how population size is related to drift processes, and it precludes the possibility that a single trait can undergo selection and drift at the same time and place. See Stephens (2004, 2007), Riesman and Forber (2005), and Barrett et al. (2012) for discussion. Pfeifer (2005) argues that selection and drift might be causally distinct if the probabilities used to define fitness abstract away from some features of the environment.

<sup>5</sup>Shanahan (1992) is one of the few dissenters from this conclusion.

## 2 A Supervenience Thesis

The claim that identical twins living in identical environments must have the same fitness follows from an appealing thesis:

- (S) The fitness of token individual  $i$  in environment  $E$  at time  $t$  supervenes on  $i$ 's other properties at time  $t$ , including the property of living in environment  $E$ .<sup>6</sup>

By hypothesis, the twins in Scriven's example have the same genotypes and phenotypes, so S entails that they must have the same fitness value *if* they live in identical environments.<sup>7</sup> If their environments are merely similar, S has no such implication.

Scriven notes that the twins have different spatial locations at the time of the lightning strike, but asserts that this doesn't count. He does not explain why. If one twin was at the top of a mountain while the other was at its foot when the lightning did its work, why doesn't that suffice to say that they differed in fitness? Otherwise similar organisms can differ in their ability to survive and reproduce because they live in different habitats. This suggests that you should stipulate, for the sake of a clean example, that the twins were both at the top (or both at the bottom) at that fateful moment. But even if they were standing side-by-side, that's still a difference (though a small one), so the question resurfaces.<sup>8,9</sup> The supervenience thesis fails to deliver the desired result, even if the twins were shoulder to shoulder. Maybe that small difference in location gave them different probabilities of being hit by lightning.

My question here is not about the truth of S but whether it entails that the twins have the same fitness. I agree that S is a sensible philosophical principle.<sup>10</sup> Since fitness is a probabilistic property, S can be motivated by thinking about such properties more generally. If two coins differ in their bias (their probabilities of landing heads when tossed), this must be because the coins are physically different, or the circumstances of their tossing are different, but do *all* probabilistic properties supervene on other properties? Maybe not; perhaps some probabilistic properties (of elementary physical particles, for instance) are brute. In any event, the thesis that biological fitness supervenes on other properties leaves it open that some of those other properties are themselves probabilistic. This opening is not a mere speculation; evolutionary game theory considers mixed strategies, meaning rules for behavior that are probabilistic (Maynard Smith 1982), and the mixed strategies that an organism follows can clearly be part of the supervenience base of its fitness.

<sup>6</sup>Note that the organism's traits at time  $t$  don't cause its fitness at time  $t$ , if cause must precede effect. Note also that this supervenience thesis permits token organisms to change their fitnesses as their environments change. This is contrary to Ramsey's (2006) claims about "block fitness."

<sup>7</sup>What matters here is whether the twins live in type-identical environments; the environments need not be token-identical.

<sup>8</sup>Brandon (1990) argues that if two organisms experience the same selection process, they must inhabit the same environment. He is talking about types, not tokens, here. Abrams (2009a, 2009b) argues that the relevant environment for considering natural selection is the one occupied by the whole population, not the sub-environments occupied by only some of the organisms in that population. Two points are in order here. First, Brandon's claim doesn't settle whether the twins differ in fitness. Second, Abrams's point seems to conflict with the fact that selection in a heterogeneous environment is an important topic in evolutionary biology (Sterelny and Kitcher 1988), which opens the door to treating the twins as occupying different sub-environments.

<sup>9</sup>Another possible objection to treating the twins as experiencing a selection process is that their presence in different sub-environments isn't heritable. My reply is that heritability is necessary for evolution by natural selection, but not for selection itself. This is clear from the breeder's equation, which separates the strength of selection from the heritability of a trait; these together determine the response to selection.

<sup>10</sup>I take no stand on whether the supervenience thesis S is part of evolutionary theory.

### 3 Why Twin-Equality-in-Fitness Isn't a Commitment of Evolutionary Theory

In their landmark paper, Mills and Beatty (1979) first define the fitness of a token individual, and then use that concept to define the fitness of a trait. Their formulation is this:

- (C) The fitness of trait  $T$  in population  $p$  at time  $t$  = the average fitness of the token individuals in population  $p$  at time  $t$  that have trait  $T$ .

Notice that proposition C coordinates trait fitness with the fitnesses of token organisms, whereas the supervenience thesis S doesn't mention trait fitnesses at all.

Coordinating principle C does not entail that the twins have the same fitness, even if they have the same genotypes and phenotypes and live in identical environments. It merely says that the conjunction of all the genetic and phenotypic<sup>11</sup> traits they have (that is, the *total trait complex* that they uniquely share) has a fitness equal to the *average* of the two individuals' fitnesses.

There is something else that proposition C does not say. It doesn't say that the concept of token-individual fitness is primary and the concept of trait fitness is derivative. Mills and Beatty don't endorse this asymmetry thesis, but others have done so.<sup>12</sup> I think C is true, but the asymmetry thesis that might be associated with it is wrong. The fitnesses of traits play a central role in evolutionary biology; the fitnesses of token organisms do no such thing.

This is fortunate, because the fitnesses of token individuals (as opposed to their *actual* longevity and their *actual* degrees of reproductive success) are almost always unknowable in practice. A token individual tastes of life but once. In that one-shot deal, it either survives to reproductive age or it does not; if it survives to reproductive age, it has some number of offspring. From this paltry data set, it would almost always be absurd to estimate the fitness of that token organism. This would be like tossing a coin just once and trying to estimate from the outcome of that one toss what its probability of heads is (Sober 2013).<sup>13</sup>

The reason I say that it is "almost always" impossible to assign a fitness to a token organism is that I want to allow for the possibility that an organism might be known from the outset to have a trait (genetic or phenotypic) that will kill the organism before it reaches reproductive age. Here there is a very good reason to say that the individual's fitness is zero. But the organisms that evolutionary biologists want to talk about are rarely like that. I also said that fitness estimates for token organisms usually aren't possible *in practice*. This leaves it open that if one were to clone a token organism a large number of times, and rear the clones in identical environments,<sup>14</sup> one could derive from that large data set a reasonable estimate of the token organism's fitness (Brandon 1990). This shows that the fitness of a token organism is knowable *in principle*. However, evolutionary biology requires more than that. It needs for *available* data to provide good estimates of fitness values. It is in that sense that the fitness of token organisms are usually empirically inaccessible.<sup>15</sup>

<sup>11</sup>Here I include the organism's environment as part of its phenotype.

<sup>12</sup>For examples, see Brandon (1978, 188), Ramsey (2006), and Pence and Ramsey (2013).

<sup>13</sup>This is not to deny that tossing a coin just once yields a maximum likelihood estimate of its probability of landing heads. The estimated probability will have a value of 1 or 0, depending on whether the coin lands heads or tails. You would be ill-advised to use the resulting estimate to predict the next toss or tosses of the same coin.

<sup>14</sup>An organism's environment includes all the other organisms (both conspecific and not) with which it interacts, so the experiment described here would require cloning those other organisms too.

<sup>15</sup>Ettinger et al. (1990, 504) say something similar. They argue that "it is only abstract entities such as types, genotypes and alleles that can be fit," and that "the fitness of an individual (should there be such a thing) cannot actually be measured or determined in any way on the basis of the individual's own actual reproductive fate; it can only be determined by taking the average success of a sample of individuals of the same type."

Although token individuals furnish data sets of size 1, traits almost always do much better. The traits that evolutionary biologists want to discuss usually are exemplified by large numbers of organisms. Each of those token organisms has just one lifetime, but averaging over those numerous lifetimes can provide a meaningful estimate of the fitness of the trait those organisms share. The analogy with coins is this: A single toss of a coin fails to furnish a good estimate of the coin's bias, but the average bias of 1,000 coins can be meaningfully estimated by tossing each coin just once. The coins may differ from each other in their biases, but that does not matter.<sup>16</sup>

With respect to the propensities of token objects, it often makes sense to discern two avenues of discovery (Sober 1984). To estimate a coin's propensity to land heads, repeated tossing is one obvious avenue, but another is to examine the coin's physical properties. If the coin is symmetrical, it seems reasonable to conclude that the coin is fair, even if you have never tossed it. However, looking at the physical make-up of the coin is not enough; you also need to examine the tossing procedure, and maybe the physical features of the tossing apparatus (a machine or a person) can be examined without your ever having to put the tossing device to work.<sup>17</sup> Even if this "design analysis" works for the coin, the fitness of a token organism is almost always far more opaque.<sup>18</sup>

The idea that an organism's fitness is just as empirically accessible as a lump of sugar's solubility has influenced many philosophers (including, at one time, me). Ayhan Sol (pers. comm.) has suggested how this seeming similarity can be mapped onto the example of the twins. Imagine two cubes of sugar. Both are water-soluble, but only one is dropped into water. The two cubes have the same propensity to dissolve, though only one of them does so. If that is the right thing to say about the sugar cubes, are we therefore obliged to say that the two twins, being physically identical, must therefore have the same propensity to survive and reproduce? I think the answer is *no*. Whether an object is water-soluble does not depend on whether it is actually dropped into water, but a token organism's fitness depends on what its actual environment is, and, as already explained, there is no uniquely correct answer to the question of whether the twins live in "the same" environment.

<sup>16</sup>Mike Steel (pers. comm.) has suggested a frequentist analysis of this comparison. Suppose you have coins 1, 2, . . . ,  $n$  with  $p_i$  being the probability that coin  $i$  lands heads. Let  $p$  be the average of these  $p_i$  values, and let  $X$  be the (random variable) proportion of heads when these  $n$  coins are each tossed once. Then the expected value of  $X$  is exactly equal to  $p$ , and the standard deviation of  $X$ ,  $SD(X)$ , is

$$SD(X) = (1/\sqrt{n})\sqrt{p - \sum_i (p_i)^2/n}.$$

Regardless of the  $p_i$  values, it follows that  $SD(X) \leq 1/(2\sqrt{n})$ , so for  $n$  large, the central limit theorem implies that the distribution of  $X$  around  $p$  will be normally distributed with a standard deviation that is small (in the case of  $n = 1000$ , at most 0.016), regardless of the actual probability values  $p_i$ . So  $X$  tells you a lot about the value of  $p$ . On the other hand, for  $n = 1$ , a single toss doesn't tell you much, since all you know about  $SD(X)$  in that case is that it is at most  $1/2$  (which it would be if  $p_1 = 1/2$ ). As long as the single coin in the  $n = 1$  experiment doesn't have a probability of heads of 1 or 0,  $SD(X)$  in the 1000 coin case is less than  $SD(X)$  in the single coin case.

A Bayesian analysis of the two coin-toss experiments is also possible. Start in both cases with a flat prior probability density distribution, then compute posterior density distributions for the 2 possible outcomes of the first experiment and the 1001 possible outcomes of the second. Then compute the distance between the prior density distribution and each possible posterior density distribution in each experiment. Finally, compare the average distance in the first experiment with the average distance in the second. The average distance in the first is a lot less than the average distance in the second.

<sup>17</sup>For discussions of the physics of coin tossing that attend to how the coin is tossed, see Keller (1986), Diaconis (1998), Diaconis et al. (2007), and Mahadevan and Yong (2011).

<sup>18</sup>Ettinger et al. (1990, 506) say that "whether or not a book binding is strong can in principle be ascertained by examining its material interactions with other bodies; the fitness of an individual organism cannot in principle be so determined."

Although I say that the fitnesses of traits and the fitnesses of token organisms are often epistemologically different, coordinating principle C describes a way in which they are connected. The fitness of a trait and the *average* fitness of the token individuals that have that trait are joined at the hip. When a trait attaches to numerous token individuals, the fitness of the trait and the average fitness of the tokens are both knowable. However, when the trait attaches to just one token individual, the fitness of the trait and the fitness of the one individual that has that trait are *usually* both unknowable. I've mentioned lethal traits as the reason I hedge my claim with the word "usually," but there is another reason.

The fitness of a target trait that happens to be possessed by just one organism can sometimes be estimated by estimating the fitnesses of other traits that are "related" to the target. For example, suppose you want to estimate the viabilities of birds that differ in the number of feathers they have. In your lab, you start with a large number of songbirds in the same species that have the same age and note which of eleven trait "bins" they are in: 2000–2099 feathers, 2100–2199, 2200–2299, ..., and 3000–3099 feathers. As you count a bird's feathers, you give it a tag that assigns the bird an identifying number and a second number that indicates how many feathers it had on the day you did your counting. Suppose you have numerous birds in each of these trait bins, except for the bin in the middle (2500–2599 feathers), which happens to contain just one bird. You then examine this bird population a year later to see which birds have died and which have survived. From this data, you can construct good estimates of the average viability (a component of fitness) of the birds in ten of the eleven bins. You graph these ten estimates (with  $x$  representing intervals of feather numbers and  $y$  representing estimated viability) and notice that they closely fit a straight line. This makes it reasonable to "interpolate"—to use that straight line to construct an estimate of the fitness of the middle trait bin in which there is just one bird.<sup>19</sup> What your interpolation provides is a reasonable estimate of the average fitness of a bird whose feather number falls in the middle bin. I hasten to add that this indirect procedure for estimating the fitness of a trait that only one organism possesses fails to generalize to the case of estimating the fitnesses of total trait complexes.

It might be suggested that the problem of assigning fitnesses to token organisms can be solved by constructing biologically plausible optimality models. My reply is that optimality models are about traits, not about token organisms. An influential example makes this clear. Hamilton (1967) developed an optimality model for sex ratio. What mix of sons and daughters should a parent produce, if the "goal" is to maximize her number of grandoffspring? Given facts about population structure and life history as inputs, Hamilton's analysis identifies an optimal sex-ratio strategy; this means that parents using the optimal strategy will, on average, be fitter than those using any other strategy in a set of alternatives. There is no commitment to the false assumption that all parents who use the same sex-ratio strategy have identical fitnesses. Still less does the model assign a numerical fitness value to a token organism. How could the model possibly do that, since an organism has many other traits that affect its fitness besides the sex-ratio strategy it happens to deploy?<sup>20</sup>

My thesis that the fitnesses of traits are usually empirically accessible while the fitnesses of token organisms rarely are may seem to fall afoul of a standard inferential procedure. Suppose

<sup>19</sup>In this example, although there is just one bird *in your data set* whose feather number falls in the middle bin, there may be other birds, not in your sample, that do the same. However, that possibility is extraneous to the point being made here. Even if there is only one bird *in the whole species* that falls in the middle trait bin, the fitness of that trait can be estimated via the indirect procedure described above.

<sup>20</sup>When a parasitic wasp has several broods in its lifetime, observing the mix of sons and daughters it produces in each brood allows you to determine how well its behavior fits an optimality model concerning sex-ratio strategy (Orzack and Sober 1994). This is a different kettle of fish from estimating the fitness of the wasp, which has many other traits.

you look at a large number of light-color juvenile moths that live on dark trees and find that only 30% of them survive to reproductive age (Kettlewell 1955). A maximum likelihood estimate of the average viability fitness of those individuals (aka the fitness of the light-coloration trait at that time and place) is 0.3. Now suppose I single out one of those moths and ask you what its fitness is. All you know about this moth is that it belongs to a set of organisms whose average viability is estimated to be 0.3. Given this information, you may think you're obliged to infer that that organism's fitness is also 0.3. However, if you're obliged to do that, you also are obliged to infer that each of the other organisms in the set has that same fitness value. This seeming obligation has landed you in the pickle of proclaiming that these organisms have exactly the same fitness value. However, this proclamation is thoroughly unrealistic, in that you know that the organisms almost certainly differ in numerous ways that affect their fitnesses.<sup>21</sup>

The claim that individuals that share a trait always have the same fitness leads to a more embarrassing conclusion; it entails that all the individuals in the whole population (not just those sharing a single trait) have the same fitness in plenty of situations where that isn't true. For example, consider a population of individuals, each of which has trait *A1* or trait *A2*, and each has trait *B1* or trait *B2*. Suppose there are organisms in each of four subgroups (*w*, *x*, *y*, *z*) in the population such that

- Individuals in *w* have *A1* and *B1*.
- Individuals in *x* have *A1* and *B2*.
- Individuals in *y* have *A2* and *B1*.
- Individuals in *z* have *A2* and *B2*.

According to the principle that individuals with the same trait must have the same fitness, individuals *within* each of these four groups are identically fit, but there is more. The principle additionally entails the individuals in *w* and *x* are equally fit (because all have *A1*), individuals in *w* and *y* are equally fit (because all have *B1*), individuals in *x* and *z* are equally fit (because all have *B2*), and individuals in *y* and *z* are equally fit (because all have *A2*), from which it follows that all the individuals in the population are equally fit. Surely no sound principle should lead to the conclusion that it is *a priori* impossible for natural selection to act on trait *A* or on trait *B* in this population. The idea that individuals that have the same trait must have the same fitness would not lead to this trouble if it were restricted to total trait complexes, but it then would not apply to claims about "singleton" traits evolving by natural selection; such claims are entirely routine in evolutionary biology.<sup>22</sup>

When I say that the fitnesses of traits are fundamental while the fitnesses of token organisms are not, I don't mean that the former are primitive and the latter are derived. Neither of these claims is true. First, the fitness of a trait *can* be defined, and evolutionary biologists and

<sup>21</sup>This point applies to the other empirical studies of selection in the wild discussed in Endler (1986); for a different take on those studies, see Brandon and Ramsey (2007).

<sup>22</sup>This point bears on Brandon's (1990) discussion of environmental homogeneity. He says that "when an environment is homogenous with respect to selection, within that environment different copies of the same 'type' will do equally well with respect to selection—that is, they will have the same relative expected reproductive success" (52). I take this comment to express a stipulative definition of what it means for an environment to be "homogenous with respect to selection," so there can be no argument about its correctness. However, Brandon uses this concept to formulate the following proposition: "Some biological entities differ in their adaptedness to their common selective homogeneous environments, this difference having its basis in differences in some traits of the entities" (149). Brandon uses "adaptedness" to refer to what I've been calling "fitness." He says that this proposition is one of two that "form the empirical biological core of the theory of natural selection" (149). It's this last quoted statement that I think the above argument calls into question.

philosophers of biology are able to do so.<sup>23</sup> In addition, the fitness of a token organism *cannot* be derived from information about the fitnesses of the traits that the individual possesses, in spite of what proposition S asserts. The fitness of a total trait complex is just as empirically inaccessible as the fitness of the token organism that has that total trait complex.<sup>24</sup>

## 4 Back to the Twins

Argument F needs replacing. Here's an improved argument for the same conclusion:

(F') Trait *T* undergoes a selection process in a population at a time precisely when trait *T* belongs to a partition of traits, each of which is instantiated in the population, and there is variation in fitness among the traits in that partition.

Drift can occur without selection, and a possible result is that individuals with trait *T*, on average, have more offspring than individuals with an alternative trait in the partition.

---

Therefore, the fitness of a trait is not the same property as the average number of offspring that individuals with that trait have.

Here I'm using the term "partition" to denote a set of traits, each of which is exemplified in the population at the time in question, and each individual has just one trait in that set.<sup>25</sup> The first premise is an agreed-on definition of natural selection. The second premise describes an agreed-on conceptual possibility. Identical twins and lightning strikes aren't needed to get the desired conclusion.

This substitute for argument F capitalizes on the fact that there can be differential reproductive success without selection, but the two concepts part ways in the other direction as well. There can be selection (aka variation in fitness) without differential reproductive success. A helpful analogy is provided by the case of two coins that differ in their probabilities of landing heads, and yet produce the same number of heads in a run of tosses. Differential reproductive success is *defeasible evidence* for the presence of natural selection, not the *proper definition* of that concept, a point that proponents of the propensity interpretation of fitness were right to

---

<sup>23</sup>Williams (1971) took fitness to be a primitive concept in evolutionary theory, and Rosenberg (1985) concurred. Denying this thesis does not commit one to holding that there is a single definable concept of fitness that works in all the many contexts in which fitness is properly applied. The point is that in each such context, the concept can be (and should be!) defined.

<sup>24</sup>After proposing a "new foundation for the propensity interpretation of fitness," Pence and Ramsey (2013) take up several objections to their proposal, one of which (objection #5) says that "the theory of evolution by natural selection fundamentally concerns trait fitnesses, not individual fitnesses." Here they discuss an earlier paper of mine (Sober 2013), in which I make the same point I've made here—that the fitnesses of token organisms are mostly unknowable—though the earlier paper does not discuss the twins. Their reply (872) is that "trait fitness is straightforwardly parasitic on individual fitness" and that "individual fitness is in some sense foundational." They do not address my reason for denying the centrality of individual (aka token) fitnesses. Pence and Ramsey's goal in their paper is to define a concept of (very) long-term fitness that takes account of "all possible future causal influences on organisms" (868); applying their proposed definition to a given individual requires that you know the individual's expected number of offspring, the expected number of offspring that each of those offspring has, and so on. This is precisely the kind of information that I claim is largely unavailable. Pence and Ramsey describe various definitions of fitness developed by biologists and claim that these are used to successfully estimate fitnesses (864, 866); in fact, the biologists cited are estimating the fitnesses of traits, not token individuals.

<sup>25</sup>I talk about trait *T* and a partition to which it belongs, rather than two alternative traits *T* and *T\**, because these two traits may be equal in fitness, even though there's variation in fitness in the partition to which these two traits belong.

emphasize. Here is yet another example in science in which operationalism gets things exactly wrong.

Besides avoiding the attribution of fitnesses to token organisms, the argument that replaces argument F has the virtue of not having to say that the twin killed by lightning was *unlucky* rather than unfit. The import of that remark is that the dead twin was *more* fit than her dismal reproductive track record suggests. But why isn't it then also true that the surviving twin was *lucky* to have the lightning bolt miss her, and so she was *less* fit than her fortunate track record suggests? Every organism is either lucky or unlucky, I guess. If we set aside the life histories of organisms that were due to luck, good or bad, our data set for estimating fitnesses, I fear, will plunge to zero. The concepts of "luck" and "accident" don't provide much help in understanding fitness and natural selection.

## 5 Concluding Comments

Scriven thinks it would be bizarre to assign the two twins different fitnesses, and I agree. However, I think it is equally unmotivated to assign them identical fitnesses. I'm doubting the presupposition shared by both assignments—that evolutionary biology is in the business of assigning fitnesses to token organisms. This point is at variance with much of the philosophical literature on fitness and natural selection, but I think it is something of a commonplace in evolutionary biology itself. When evolutionary biologists assign fitnesses to phenotypes or genotypes, they are talking about the fitnesses of traits that typically are present in multiple individuals. Though biologists rarely use the philosophical jargon of "type" and "token," their terminology of *phenotypes* and *genotypes* is apt; they aren't assigning fitnesses to *phenotokens* or *genotokens*.

The idea that the fitnesses of traits are fundamental to the theory of evolution leaves open what sorts of traits have fitness values and what the objects are that have those traits. As mentioned, the traits can be both phenotypic and genetic. Being an *AA* homozygote is just as much a property of organisms as running fast is. In this respect, I see a continuity between Darwin's theorizing, theorizing in the Modern Synthesis, and recent evolutionary biology as well. Evolutionary theory has always been about traits. This is not to deny, of course, that the list of traits that evolutionary biologists address has greatly expanded.

Traits cause survival and reproductive success only in virtue of their attaching to objects. This is a truism about causality: curiosity can kill the cat only because some object (e.g., the cat) is curious. Darwin talked about whole organisms having traits that affect their ability to survive and reproduce, and the Modern Synthesis remained faithful to this formula. Fisher (1930) talked about genes, but he did so in order to talk about the genetic properties of whole organisms. In discussing heterozygote superiority, for example, he was talking about organisms being heterozygous, and about the average fitness of the organisms that have a given heterozygote genotype. Indeed, whole organisms figure indispensably in his conception of gene frequencies in populations; the frequency of a gene is computed by finding either zero, one, or two copies of the gene at a given locus *in each organism* in the population (when the organisms are diploid). Fat organisms count no more than thin ones, even though the former have more cells, and therefore more gene tokens.<sup>26</sup>

With respect to groups, there is an important difference between Darwin and Fisher. Darwin (1859, 1871) made room for group selection in his theory (Sober 2011), and so he talked about groups differing in their fitnesses. R. A. Fisher (1930), on the other hand, had little to say

<sup>26</sup>See Sober (1984) for discussion of this "head-counting paradigm" (29–30) and the related idea that organisms are often treated as "benchmarks" of selection (356).

about that possibility, and most of what he said was negative. Fisher didn't deny that groups exist; indeed, populations are groups, and selection occurs in them. But he denied that one needs to think about the fitnesses of groups. For Fisher, there is selection *in* groups, not selection *of* groups.

More recently, the subject of intragenomic conflict has interested evolutionary biologists; it was unknown to Darwin and to the early architects of the Modern Synthesis as well. Meiotic drive is a familiar example of intragenomic conflict; in discussing its evolutionary consequences, one needs to think about genes (or chromosomes) differing in their contributions to the gamete pool. In simple models of this process, whole organisms have one of the three genotypes *ww*, *dw*, and *dd* (where *w* is the wild type and *d* is a driving gene), but talking about the fitnesses of organisms is not enough. One also needs to look inside of organisms and describe how their genes differ in fitness. Organisms that are *ww* and organisms that are *dw* may be equally fit (on average), but the *d* gene is fitter than the *w* gene in *dw* heterozygotes (Sober and Wilson 1998, 89–90). In those heterozygotes, chromosomes with a copy of gene *d* are on average fitter than chromosomes with a copy of gene *w*.

The continuity I have described between Darwin and Fisher is at odds with Denis Walsh's (2015) view of the two. After quoting Fisher's (1930) famous analogy between his fundamental theorem of natural selection and the second law of thermodynamics, Walsh says that Fisher's analogy embodies

population thinking in its most rarified form. It involves 'ignoring individuals' in a comprehensive and radical way. Individual organisms are no part of the ontology of Fisher's theory, nor do they figure in its explanatory apparatus. (56)<sup>27</sup>

On the next page, Walsh adds that

it is extremely important to recognize the degree to which Fisher's statistical approach to population change is a departure from Darwin's causal account of population change. In a certain sense, they aren't even about the same kind of thing. Darwin's theory plots the change in frequency of *lineages of organisms* in a population, as a function of their success in the struggle for life. Fisher's theory plots the change in *intrinsic growth rate* of an indefinitely large population of *abstract "gene ratios"* as a function of its statistical distribution of growth rates. They don't appear to have a whole lot in common. (57)

Here Walsh embraces one of the four theses that Walsh, Ariew, and Matthen (2017) defend under the heading of *statisticalism*, namely the claim that Fisher's models of natural selection (and others in the Modern Synthesis) aren't causal, whereas Darwin's theory is.

What is true, I think, is that Darwin talked about the phenotypes of organisms, whereas Fisher talked about their genotypes. This *is* a profound shift. Fisher helped invent population genetics; Darwin, of course, did no such thing. However, both embraced population thinking by creating theoretical frameworks in which fitness differences among traits explain changes in the frequencies of traits in populations, and both did so by thinking about the consequences that traits have for the survival and reproduction of organisms.<sup>28</sup> The question of whether Darwin's theory and Fisher's are both *causal*<sup>29</sup> is separable from the question of whether each is about *traits* and separable also from the question of whether each includes whole organisms in its ontology.

<sup>27</sup>Here, Walsh is using Mayr's (1959) influential idea of "population thinking"; for discussion, see Sober (1980).

<sup>28</sup>Darwin introduced the word "fitness" into the *Origin of Species* only in its fifth edition, but that doesn't mean that he wasn't talking about fitness all along.

<sup>29</sup>For discussion of the causal question, see Shapiro and Sober (2007).

## Acknowledgments

I am grateful to John Beatty, Robert Brandon, Katie Deaven, Mohan Matthen, Emi Okayasu, Steven Orzack, Ayhan Sol, Mike Steel, Christopher Stephens, Joel Velasco, and Denis Walsh for their help.

## Literature cited

- Abrams, Marshall. 2009a. "Fitness 'Kinematics': Biological Function, Altruism, and Organism–Environment Development." *Biology and Philosophy* 24: 487–504.
- Abrams, Marshall. 2009b. "What Determines Biological Fitness? The Problem of the Reference Environment." *Synthese* 166: 21–40.
- Barrett, Martin, Hayley Clatterbuck, Michael Goldsby, Casey Helgeson, Brian McLoone, Trevor Pearce, Elliott Sober, Reuben Stern, and Naftali Weinberger. 2012. "Puzzles for ZFEL, McShea and Brandon's Zero Force Evolutionary Law." *Biology and Philosophy* 27: 723–735.
- Beatty, John. 1980. "Optimal-Design Models and the Strategy of Model Building in Evolutionary Biology." *Philosophy of Science* 47(4): 532–561.
- Beatty, John. 1984. "Chance and Natural Selection." *Philosophy of Science* 51: 183–211.
- Beatty, John, and Susan Finsen. 1989. "Rethinking the Propensity Interpretation: A Peek inside Pandora's Box." In *What the Philosophy of Biology Is*, edited by Michael Ruse, 17–30. Dordrecht: Kluwer.
- Brandon, Robert. 1978. "Adaptation and Evolutionary Theory." *Studies in History and Philosophy of Science* 9(3): 181–206.
- Brandon, Robert. 1990. *Adaptation and Environment*. Princeton: Princeton University Press.
- Brandon, Robert. 2005. "The Difference between Drift and Selection: A Reply to Millstein." *Biology and Philosophy* 20: 153–170.
- Brandon, Robert, and Grant Ramsey. 2007. "What's Wrong with the Emergentist Statisticalist Interpretation of Natural Selection and Random Drift?" In *Cambridge Companion to the Philosophy of Biology*, edited by David Hull and Michael Ruse, 281–303. Cambridge: Cambridge University Press.
- Casselman, Anne. 2008. "Identical Twins' Genes are not Identical." *Scientific American*, April 3.
- Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection*. London: Murray.
- Darwin, Charles. 1871. *The Descent of Man and Selection in Relation to Sex*. London: Murray.
- Diaconis, Percy. 1998. "A Place for Philosophy? The Rise of Modeling in Statistical Science." *Quarterly of Applied Mathematics* 56: 797–805.
- Diaconis, Percy, Susan Holmes, and Richard Montgomery. 2007. "Dynamical Bias in the Coin Toss." *Society for Industrial and Applied Mathematics (SIAM) Review* 49: 211.
- Drouet, Isabelle, and Francesca Merlin. 2015. "The Propensity Interpretation of Fitness and the Propensity Interpretation of Probability." *Erkenntnis* 80: 457–468.
- Endler, John. 1986. *Natural Selection in the Wild*. Princeton: Princeton University Press.
- Ettinger, Lia, Eva Jablonka, and Peter McLaughlin. 1990. "On the Adaptations of Organisms and the Fitness of Types." *Philosophy of Science* 57: 499–513.
- Fisher, Ronald. A. 1930. *The Genetical Theory of Natural Selection*. New York: Dover Books, 1957.
- Frank, Steven, and Montgomery Slatkin. 1990. "Evolution in a Variable Environment." *The American Naturalist* 136(2): 244–260.

- Gillespie, John. 1977. "Natural Selection for Variances in Offspring Numbers—A New Evolutionary Principle." *American Naturalist* 111: 1010–1014.
- Hamilton, William 1967. "Extraordinary Sex Ratios." *Science* 156(3774): 477–488.
- Hodge, Jonathan. 1987. "Natural Selection as a Causal, Empirical, and Probabilistic Theory." In *The Probabilistic Revolution*, volume 2, edited Lorenz Krüger, Gerd Gigerenzer, and Mary Morgan, 233–279. Cambridge: MIT Press.
- Hull, David. 1974. *Philosophy of Biology*. Englewood Cliffs: Prentice-Hall.
- Keller, Joseph. 1986. "The Probability of Heads." *American Mathematical Monthly* 93: 191.
- Kettlewell, Henry. 1955. "Selection Experiments on Industrial Melanism in the Lepidoptera." *Heredity* 9(3): 323–342.
- Kimbrough, Steven. 1980. "The Concepts of Fitness and Selection in Evolutionary Biology." *Journal of Social Biological Structures* 3: 149–70.
- Mahadevan, Lakshminarayanan, and Ee Hou Yong. 2011. "Probability, Physics, and the Coin Toss." *Physics Today*, July, 66–67.
- Matthen, Mohan, and André Ariew. 2002. "Two Ways of Thinking about Fitness and Natural Selection." *Journal of Philosophy* 99: 55–83.
- Mayr, Ernst. 1959. "Typological versus Population Thinking." In *Evolution and Anthropology: A Centennial Appraisal*, 409–412. Washington: The Anthropological Society of Washington. Reprinted in Mayr (1976), *Evolution and the Diversity of Life*. Cambridge: Harvard University Press.
- Maynard Smith, John. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- Michod, Richard. 2000. *Darwinian Dynamics*. Princeton: Princeton University Press.
- Mills, Susan, and John Beatty. 1979. "The Propensity Interpretation of Fitness." *Philosophy of Science* 46: 263–86.
- Millstein, Roberta. 2002. "Are Random Drift and Natural Selection Conceptually Distinct?" *Biology and Philosophy* 171: 33–53.
- Millstein, Roberta. 2016. "Probability in Biology: The Case of Fitness." In *The Oxford Handbook of Probability and Philosophy*, edited by Alan Hájek and Christopher Hitchcock, 601–622. Oxford: Oxford University Press.
- Millstein, Roberta. 2017. "Genetic Drift." In *The Stanford Encyclopedia of Philosophy*, edited by Edward Zalta. <https://plato.stanford.edu/archives/fall2017/entries/genetic-drift/>.
- Orzack, Steven, and Elliott Sober. 1994. "Optimality Models and the Test of Adaptationism." *American Naturalist* 143(3): 361–380.
- Pence, Charles, and Grant Ramsey. 2013. "A New Foundation for the Propensity Interpretation of Fitness." *British Journal for the Philosophy of Science* 64: 851–881.
- Pfeifer, Jessica. 2005. "Why Selection and Drift Might be Distinct." *Philosophy of Science* 72: 1135–1145.
- Ramsey, Grant. 2006. "Block Fitness." *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 37: 484–498.
- Reisman, Kenneth, and Patrick Forber. 2005. "Manipulation and the Causes of Evolution." *Philosophy of Science* 72: 1115–1125.
- Rosenberg, Alexander. 1985. *The Structure of Biological Science*. Cambridge: Cambridge University Press.

- Scriven, Michael. 1959. "Explanation and Prediction in Evolutionary Theory." *Science* 130: 477–82.
- Shanahan, Timothy. 1992. "Selection, Drift, and the Aims of Evolutionary Theory." In *Tree of Life: Essays in Philosophy of Biology*, edited by Paul Griffiths, 131–61. Dordrecht: Kluwer.
- Shapiro, Lawrence, and Elliott Sober. 2007. "Epiphenomenalism—The Do's and the Don'ts." In *Thinking about Causes*, edited by Peter Machamer and Gereon Wolters, 235–264. Pittsburgh: University of Pittsburgh Press.
- Sober, Elliott. 1980. "Evolution, Population Thinking, and Essentialism." *Philosophy of Science* 47(3): 350–383.
- Sober, Elliott. 1984. *The Nature of Selection*. Cambridge: MIT Press.
- Sober, Elliott. 2001. "The Two Faces of Fitness." In *Thinking about Evolution: Historical, Philosophical, and Political Perspectives* volume 2, edited by Rama Singh, Diane Paul, Costas Krimbas, and John Beatty, 309–321. Cambridge: Cambridge University Press.
- Sober, Elliott. 2011. *Did Darwin Write the Origin Backwards?* Amherst: Prometheus Books.
- Sober, Elliott. 2013. "Trait Fitness is not a Propensity, but Fitness Variation Is." *Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 336–341.
- Sober, Elliott, and David Sloan Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Cambridge: Harvard University Press.
- Stephens, Christopher. 2004. "Selection, Drift and the 'Forces' of Evolution." *Philosophy of Science* 71(4): 550–570.
- Stephens, Christopher. 2010. "Forces and Causes in Evolutionary Theory." *Philosophy of Science* 77(5): 716–727.
- Sterelny, Kim, and Philip Kitcher. 1988. "The Return of the Gene." *Journal of Philosophy* 85(7): 339–361.
- Triviño, Vanessa, and Laura Nuño de la Rosa. 2016. "A Causal Dispositional Account of Fitness." *History and Philosophy of Life Sciences* 38: 1–18.
- Walsh, Denis. 2015. *Organisms, Agency, and Evolution*. Cambridge: Cambridge University Press.
- Walsh, Denis, André Ariew, and Mohan Mathhen. 2017. "Four Pillars of Statisticalism." *Philosophy, Theory, and Practice in Biology* 9: 1–18.
- Williams, Mary. 1971. "Deducing the Consequences of Evolution: A Mathematical Model." *Journal of Theoretical Biology* 29(3): 343–385.

---

© 2020 Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits anyone to download, copy, distribute, display, or adapt the text without asking for permission, provided that the creator(s) are given full credit.

ISSN 2475-3025