

Not A Sure Thing: Fitness, Probability and Causation
Denis Walsh (Toronto)

Abstract: This is a defence of the statistical interpretation of fitness. I distinguish three conceptions of fitness by differentiating three interpretations of the evolutionary explanations in which fitness figures. These are: the Two Factor Model, The Single Factor Model and the Statistical Interpretation. These interpretations differ in their degrees of causal commitment. The first two are committed to fitness distribution being a cause of population change. The last maintains that fitness correlates with, but does not cause population change. The defence of the Statistical Interpretation relies upon a peculiar feature of fitness pointed out by John Gillespie. The Gillespie conception of fitness demonstrates that fitness distribution cannot be a cause of population change.

Fitness is the central explanatory concept in evolutionary theory. Despite its importance, however, there is little agreement on what fitness is. There is some consensus to be sure: (i) Natural selection occurs when there is variation in trait fitness in a population; (ii) The distribution of trait fitnesses predicts and explains changes population structure; (iii) Trait fitness is a function of the mean and variance of the expected reproductive output of individuals with the trait. But beyond this meagre set of shared commitments there is little common ground. The majority opinion is that fitness is a causal property. Fitness is (or measures) the propensity of a trait type to change in relative frequency in a population. Accordingly, fitness distribution is a causal propensity of a population: its tendency to undergo selective change.¹ Natural selection, it seems to follow, is a population-level causal process; it is that process caused, and measured, by fitness

¹ By 'fitness distribution' I mean all the trait fitnesses in a population taken together.

Not A Sure Thing

distribution. Natural selection explanations, then, are causal explanations in that they cite causes of population change.

I find this majority view implausible. I offer support for an alternative interpretation of fitness and the evolutionary explanations in which it figures. This Statistical Interpretation was proposed in outline by Matthen and Ariew (2002) and Walsh et al (2002) and has been the subject of a considerable amount of comment—mostly negative—since. It holds that fitness is a mere statistical, non-causal property of trait types, so fitness distribution is a statistical, non-causal property of a population. Fitness distribution explains, but does not cause, the changes in a population undergoing natural selection. Evolutionary explanations—those that cite fitness distributions—are ‘mere statistical’ non-causal explanations.

My defence of the Statistical Interpretation proceeds in the following way. I survey three current, competing interpretations of natural selection explanations—the ‘Two Factor Model’, the ‘Single Factor Model’ and the ‘Statistical Interpretation’—each of which offers a distinctive account of fitness. These interpretations represent a nested hierarchy of decreasing causal commitment; each one in the list takes on the causal commitments of its successor, plus some extra. My argument involves simply stripping away layers of excess causal commitment. This process of metaphysical divestment terminates at the Statistical Interpretation. The principal commitment shared by the first two interpretations, and shed by the Statistical Interpretation, is that fitness is a causal property. My argument relies upon a particular conception of fitness, originating with John Gillespie (1973, 1974, 1979).

1. Gillespie Fitness

Trait fitness is a statistical property of trait types. The significance of this feature of fitness is underscored by the groundbreaking work of John Gillespie (1973, 1974, 1977). Gillespie starts with the unobjectionable observation that any factor that systematically affects the propensity of a trait to change its relative frequency in a population ought to be factored into that trait's fitness. Typically, the fitness of a trait is taken to be the mean of the individual fitnesses of the organism with that trait.

...whatever selective advantage a_2 possesses over a_1 is manifested in the kinds of phenotypes these alleles or genotypes, respectively, determine. If in a population the individuals carrying a_2 are, *on average*, more viable, or longer lived, ... the former will leave more offspring than the latter and the frequency of a_2 will consequently increase in the next generation. (Grant 1963 p. 193. Quoted in Gillespie 1977: 1011. Emphasis in Gillespie 1977).

This is a commonplace idea in evolutionary texts. It fosters a very intuitively appealing picture of the explanatory role of fitness. The trait structure of a population changes as a function of the causal capacities of individual organisms (their 'ecological fitness' (Rosenberg 2006)). A trait's tendency to change in a population (its fitness) is inherited from the capacities of those individuals that possess the trait. Trait fitness is a sort of summation, or generalization, of the causal contributions of a trait type to the survival and reproduction of the individuals that possess it (Brandon and Beatty 1984, Beatty 1992, Beatty and Finsen 1989, Haug 2007).

Gillespie demonstrates that mean individual fitness is not always the best predictor of change in frequency. Variance counts too. If two traits $Trait_1$ and $Trait_2$ have

Not A Sure Thing

the same mean reproductive output but differ in their variance their frequencies will change in a predictably different way. Variance makes two sorts of contribution to fitness. When there is variance in reproductive output of individuals with a given trait between generations, the “best measure of fitness turns out to be the geometric mean of offspring number, averaged over time...” (Gillespie 1977: 1011). The fitness of a trait (w_i) is measured in the following way:

$$w_i = \mu - \frac{1}{2}\sigma_i^2$$

(where μ_i is mean reproductive output of i and σ_i^2 is the variance in reproductive output of i between generations.) Philosophers have been quick to take on board some of the implications of temporal variation in reproductive output (Beatty and Finsen 1989, Sober 2001). Yet there is another way in which variance contributes to fitness whose implications, I believe, have not been adequately explored. Gillespie tells us that where there is variation within generations for reproductive output, the fitness of a trait is to be calculated as:

$$(1) \quad w_i = \mu_i - \sigma_i^2/n$$

(where μ_i is mean reproductive output of i and σ_i^2 is the variance in reproductive output of i within a generation and n is the population size).² There are two salient features of this measure of fitness. The first, as discussed, is that variance affects fitness. Increasing variance reduces fitness. The second is that population size affects fitness. Decreasing population size reduces fitness. These two effects interact; the effect of variance on

² Sober (2001) sees the spectre of within generation variance. He suggests that it prevents us from interpreting fitness as an intrinsic property of a trait type. Abrams (2007a) also discusses some of the implications of this measure of fitness. See also Rosenberg (2006)

Not A Sure Thing

fitness decreases as an inverse function of population size. As population size increases, w_i converges on μ_i . This interaction forms an important part of the case against any causal interpretation of fitness. For now, however, one uncontroversial implication of Gillespie's work is that trait fitness—whatever else it may be—is a statistical parameter, a function of the mean and variance of reproductive output.

2. Two Factor Model

Any interpretation of evolutionary explanations must give a satisfactory account of the relation between selection and drift. Selection and drift are discrete, discernible, complementary effects. Selection is the expected population change given the fitness distribution; drift is deviation from expectation. They are independent; selection can occur without drift, drift without selection, or the two can occur in combination. The Two Factor Model is motivated by an attempt to understand this relation. Its proponents argue that we can distinguish between selection and drift *as effects* because they are respectively the consequences of two discrete, composable, proprietary *processes* (also known as 'selection' and 'drift') (Sober 1984, Millstein 2002, Stevens 2004, Shapiro and Sober 2007, Sober 2008: 195*f*). They are discrete in the sense that the conditions for selection to act in a population are wholly different, and independent, from the conditions required for drift (Brandon 2005). They are composable in the sense that population change is the consequence of the distinct contributions of selection and drift combined. Finally, selection and drift are proprietary in the sense that what it is for a population

Not A Sure Thing

change to be *selection the effect* is simply for it to be caused by *selection the process*, similarly, *mutatis mutandis*, for drift.

Sober (1984) introduces an analogy between selection and drift on one hand and forces in classical mechanics on the other that illustrates the salient features of the Two Factor Model. Newtonian forces are the paradigm of discrete, composable causes. The net force acting on a body is the sum of the distinct forces acting severally, each of which could act independently of the others. Newtonian mechanics offers a way of decomposing this net cause into component causes. While this Newtonian analogy has drawn extensive criticism—selection and drift are not literally forces—it retains a certain heuristic value. Selection and drift, like Newtonian forces, appear to be discrete and composable.³ Stephens defends the Newtonian analogy in the following way:

... evolutionary theory is analogous to Newtonian mechanics in many ways. In particular it makes perfect sense to think of selection, mutation, migration and drift as causes since they are factors that *make a difference*... Furthermore these causal factors can often combine in Newtonian ways, with one factor canceling out or augmenting the effect of another (2004: 568. Emphasis in original).

Support for the analogy comes from the observation that the putative processes of selection and drift are distinct difference makers. Each bears a different invariance relation to population change. The amount of selection in a population varies as a

³ They are not proprietary causes in that there are not special categories of, say, electromagnetic acceleration and gravitational acceleration. There is no reason in principle why mechanical effects could not be distinguished in this way, in an analogous fashion to the way that we recognise kinds of change in trait frequencies—selection and drift.

Not A Sure Thing

function of the degree of variation in fitness. The amount of drift varies as a function of population size. (Stephens 2004)).

We view selection and drift as distinct processes whose magnitudes are represented by distinct population parameters (fitnesses on the one hand, effective population size on the other). (Shapiro and Sober 2007: 261)

So, it looks like we have two independently measurable, composable causes of population change: selection and drift. This is the crux of the Two Factor Model. It is a simple, compelling idea and it has considerable currency amongst philosophers of biology (Stephens (2004), Millstein (2006), Abrams (2007b), Shapiro and Sober (2007), Reisman and Forber (2005)).

Recent work on interventionist approaches to causation appears to bear the Two Factor Model out. On the interventionist approach, manipulability is the mark of a cause (Woodward 2003). A factor, X , is a genuine cause of some effect, Z , only if an intervention that changes the value of X (i.e. manipulating it) would bring about a systematic change in the value of Z . A two factor causal model might be depicted in the following way:

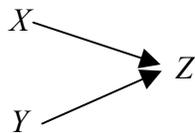


Fig 1. Diagram of a two factor causal model.

In Figure 1, X and Y are causes of Z just if there is a change relating invariance relation between X and Z and a different invariance relation between Y and Z and interventions on

Not A Sure Thing

X and Y each brings about a change in the value of Z (Woodward 2003). Where X and Y are probabilistic causes the following equation expresses the causal relations of the system:

$$(2) \quad Z = aX + bY + U$$

aX and bY represent the functional relation between X and Z and Y and Z respectively; they give us the expected values of Z given the values of X and Y . U is the error term.

Reisman and Forber (2005) argue that this sort of model fits the relation between selection (X), drift (Y) and population change (Z) perfectly. Manipulating fitness distribution alters the expected outcome of population change. Manipulating population size changes the likelihood of the outcome of selection diverging from expectation.

Natural selection occurs when ... different types of variants have different rates of survival or reproduction. This means that we can manipulate which types are favored by selection or how strongly selection favors some type over others by manipulating those factors ... that influence the expected rates of survival and reproduction. (Reisman and Forber 2005:1115)

Drift occurs when there are fluctuations in survival or reproduction due to contingent environmental events or finite population size. ... the smaller the population, the stronger the fluctuations. This means that we can manipulate the strength of drift in a population by manipulating the size of the population.

(Reisman and Forber 2005: 1115)

They illustrate this claim with a series of experiments performed by Dobzhansky and Pavlovsky (1957) in which population size is manipulated. Reisman and Forber show that these interventions have distinct, discernible effects on population change, precisely the

Not A Sure Thing

effects predicted by the two factor causal model. Manipulating fitness distributions is also a commonplace evolutionary experiment. It too leads to precisely the sort of differences in population change that the Two Factor Model predicts. Shapiro and Sober strongly endorse the argument from manipulability.

If you intervene on fitness values while holding fixed population size, this will be associated with a change in the probability of different trait frequencies in the next generation. And the same is true if you intervene on population size and hold fixed the fitnesses. (Shapiro and Sober 2007: 261)

The Two Factor Model, then, looks to be on firm footing, but it has two significant flaws. The first is that it misapplies the interventionist criterion for causes. The interventionist approach to causation does not support the claim that selection and drift are distinct causes. The second is that it fails in its principal objective, to capture the required distinction between selection and drift *as effects*. The Gillespie conception of fitness brings both of these deficiencies into focus.

Demonstrating an invariance relation between X and Z on one hand and Y and Z on the other is not sufficient to confirm the Two Factor Model depicted in Fig 1. These relations also need to be modular. Woodward explains the modularity requirement in the following way.

The basic idea that I want to defend is that the components of a mechanism should be independent in the sense that it should be possible in principle to intervene to change or interfere with the behavior of one component without necessarily interfering with the behavior of others. (Woodward 2002:S374)

Not A Sure Thing

If we make the ... plausible assumption that a necessary condition for two mechanisms to be distinct is that it be possible (in principle) to interfere with the operation of one without interfering with the operation of the other and vice versa, we have a justification for requiring that systems of equations that fully and correctly represent a causal structure should be modular. (Woodward 2003: 48)

Formally, modularity is a property of a system of equations (Woodward 2003). The equation that describes the Two Factor Model in Figure 1

$$(2) \quad Z = aX + bY + U$$

is modular if and only if we can intervene on the value of Y without altering the value of X and, conversely, we can intervene on the value of X without altering the value of Y . X and Y are discrete causes of the value of Z only if (2) is modular.

The Gillespie conception of fitness demonstrates that where X is fitness distribution (the putative cause of selection) and Y is population size (the putative cause of drift), the system of equations (2) is not modular. The reason is fairly obvious. Where there is within generation variation in reproductive output. Fitness is calculated as per

$$(1) \quad w_i = \mu_i - \sigma_i^2/n.$$

Here, population size is a component of fitness. Intervening on population size (Y) affects fitness distribution (X). A simple example illustrates the point. Equation (1) entails that the effect of variance, σ_i^2 , on fitness decreases with increasing population size, n . Given two traits with the same mean reproductive output (μ_i), but different variance, their fitnesses will diverge with decreasing population size and converge with increasing population size. This is no mere aberrant case; wherever there is variation amongst the w_i s, intervening on population size will change the distribution of fitnesses. Thus, an

Not A Sure Thing

intervention on population size is *eo ipso* an intervention on fitness distribution. Equation (2) is not modular.⁴ The relation between fitness distribution (the putative mechanism of selection) population size (the putative mechanism of drift) cannot be as depicted in Figure 1. Selection and drift are not discrete causes of population change as the Two Factor Model contends.

Clearly the experiments cited by Reisman and Forber (2005) demonstrate that manipulating population size has consequences for the kind and degree of population change. Manipulating fitness distribution also has distinct consequences for population change. So there is some reason to believe that these are population-level causes.⁵ However, the experiments do not show that fitness distribution and population size can be manipulated independently. Thus the experiments cited by Reisman and Forber do not support the Two Factor Model.

The failure of modularity has two sorts of adverse implications for the Two Factor Model. First, it demonstrates that the Two Factor Model fails on its own terms. It fails to represent selection and drift as discrete, independent causes of population change. They are not discrete because the putative mechanism of drift (population size) is a determinant of the mechanism of selection (fitness distribution). They are not independent in the sense that wherever there is variation in trait fitnesses, one cannot intervene on drift without also affecting the process of selection.

⁴ There is another significant failure of the Two Factor Model. Where the putative causal relations are expressed in (2) $Z=aX + bY + U$, the equation fails the requirement of level invariance (Woodward 2003:322, 329). An intervention on either X or Y alters the error structure U .

⁵ Although an argument by Walsh (2007) suggests that they are not.

Not A Sure Thing

Second, the Two Factor Model also fails to fulfil the principal desideratum of *any* interpretation of evolutionary explanations. It fails to distinguish between selection and drift *as effects*. It attempts to do so by representing each as the exclusive consequence of a distinct proprietary cause. But, as characterized by the Two Factor Model, selection and drift cannot be respectively the proprietary causes of selection and drift the effects. For example, a change in population structure that is precisely concordant with the expectation generated by fitness distribution ought to count exclusively as selection (the effect)—there is no drift. But, according to the Two Factor Model this outcome is generated jointly by the mechanisms of selection (fitness distribution) and drift (population size). The outcome counterfactually depends upon each in a distinct way. This raises a dilemma for the Two Factor Model; either selection the effect is jointly caused by selection the process and drift the process, or this population change does not count exclusively as selection the effect (in fact no population change does). On the first horn, selection and drift fail to be *proprietary* causes of population change, as the Two Factor Model requires. But to concede this is to give up the very idea that motivates the Two Factor Model in the first place. On the second, selection and drift cannot be distinguished *as effects*. But to concede that is to admit to the inadequacy of the Two Factor Model to meet even the most basic requirement of any interpretation of evolutionary theory.

There is a simpler, more obvious objection to the Two Factor Model; it takes on unwarranted, supernumerary causal commitments. On any causal interpretation, the relation between fitness distribution, however it is characterized, and population change is a probabilistic one. In the above model it is written as:

Not A Sure Thing

$$(2) \quad Z = aX + bY + U$$

where aX is the functional relation between fitness distribution and population change, Z , and U is the error term. The concept of drift was introduced into evolutionary theory precisely to play the role of the error term (Wright 1931). If drift is to have a role in this model, then priority and common usage require that it be assigned the role of the error term, U . Whatever other, distinct causal factors there may be in population change, they are not drift.

The same considerations apply equally to any variant of the Two Factor Model. Millstein (2002, 2006), for example, proposes that there are two distinct kinds of population-level causal processes, discriminate sampling and indiscriminate sampling.⁶ Discriminate sampling—selection—occurs when individuals with variant trait types in a population vary in their propensity to be ‘sampled’ (i.e. to survive or reproduce). Indiscriminate sampling—drift—occurs when individuals of different trait types do not differ in their propensity to be sampled. Discriminate sampling is a probabilistic cause, on this view, so selection is a probabilistic cause. The equation that describes the relation between selection and population change needs an error term. The traditional role assigned to drift is that of the error term. So, even if there is a separate process of indiscriminate sampling, it is not drift (Brandon 2005).

The distinguishing feature of the Two Factor Model is that it represents drift as a distinct cause of population change, independent from selection. But this is a

⁶ The distinction between selection and drift as (respectively) discriminate and indiscriminate sampling processes seems to have originated with Beatty (1984) and has been elaborated by Hodge (1987)

misconstrual. Drift was only ever intended as error. In a probabilistic causal model, error is not an additional cause.⁷ The Two Factor Model succeeds in representing drift implicitly *by* representing selection as a probabilistic cause with an error term. However, it fails correctly to identify drift *as error*. Then, to pile error on error (so to speak), it posits a supernumerary causal factor and calls that ‘drift’.

3. Single Factor Model

If the Two Factor Model trips over its excess metaphysical baggage, the Single Factor Model offers to lighten the load. It preserves the core commitment of any causal interpretation of evolutionary theory, that selection is a cause of population change: “Selection is the causal concept *par excellence*. Selection for properties causes differences in survival and reproductive success.” (1984: 100). On the Single Factor Model fitness distribution is a probabilistic propensity of a population (Ramsey and Brandon 2007); it predicts and explains the changes in a population undergoing selection, but drift is not a separate cause; it is just error.

Drift is any deviation from the expected levels due to sampling error. Selection is differential survival and reproduction that is due to ... expected differences in reproductive success. (Brandon 2005: 168-9)

⁷ To be fair, while Wright (1931) does introduce drift as the result ‘random sampling’, he often (e.g. 1948) treats it as though it were an independent factor explained and predicted by population size, in much the way that the Two Factor Model does.

Not A Sure Thing

Certainly, the Single Factor Model preserves the intended usage of the concept of drift, and arguably of the concept of selection. Selection is generally construed as a cause of population change. The Single Factor Model has other virtues too; another analogy serves to highlight them.

3.1 The Regression Analogy

Where the Two Factor Model explicitly draws an analogy with Newtonian mechanics, the Single Factor Model suggests another, more germane analogy for the relation between trait fitness distribution, drift and population change. The relation between population change and fitness distribution is like the relation between the ‘response’ and ‘explanatory’ variables in a linear regression. A linear regression equation, of the form

$$(3) \quad y = ax + b$$

describes a line through the scatter of points, that plot the values of two variables, y and x against each other. In (3), y is the response variable x is the explanatory variable, a describes the slope of the line, and b the y -intercept. (Hereafter, I shall assume that b passes through the origin, and drop the b term). The regression line minimizes the sum of the errors (squared) along the y axis. It is the line that best represents the dependence of y on x .

For any given point, i , in the scatter plot, the y -value can be thought of as composed of two values:

$$(4) \quad y_i = ax_i + \varepsilon_i$$

where ax_i is the expected value generated by the linear relation between y and x and ε_i is the divergence of y_i from then expected value, error. The relations between change in a

Not A Sure Thing

given population, fitness distribution and drift are analogous to those between y_i , ax_i and ε_i respectively. The ax_i term represents the effect of fitness distribution on population change, ‘selection’ in a word. This relation generates an expected outcome for population change. The difference between this expected outcome and the observed, y_i , is ε_i —drift.

The analogy captures certain important features of the relation between selection and drift, in a way that the Two Factor Model signally failed to do. Under certain plausible assumptions, the functional relation in a regression, ax_i , makes no prediction about the direction or magnitude of ε_i (and of course the converse relation holds). The expected outcome, ax_i , and error, ε_i , are thus independent in the same way that selection and drift are said to be. The regression analogy also captures a further important feature of drift. As noted, there is a predictable relation between the magnitude of drift and population size. Drift is larger in small populations.⁸ This is to be expected if drift is the sort of statistical error we find in regression analyses. Statistical error increases with decreasing sample size. But there is no temptation to think of this relation as causing a particular error value. When an individual value, y_i , deviates from the expected outcome, ax_i , one does not think of the error, ε_i , as being caused by the sample size.

It is tempting to think of the equation that describes the value of y_i in a regression relation

$$(3) \quad y_i = ax_i + \varepsilon_i$$

as an instance of the causal relation

$$(2) \quad Z = aX + U.$$

⁸ This is the relation, recall, that misled the Two Factor Model into thinking that drift is a causal mechanism.

Not A Sure Thing

This is the essence of the Single Factor Model. Fitness distribution is a probabilistic cause of population change. Drift is the error term.

3.2 Interpreting the Analogy

The regression analogy is clearly congenial to the Single Factor Model of evolutionary theory, but it does not particularly support the Single Factor Model. The reason is that a regression relation is not a causal relation. It is a statistical correlation. Sometimes the correlation between explanatory (x) and response (y) variables is a consequence of x being a cause of y and sometimes it isn't. The regression relation between plant height and sunlight, for example, clearly reflects the efficacy of sunlight on plant growth. But, an old chestnut from introductory statistics classes illustrates the well-known perils of inferring causes directly from regression relations. It is said that a regression relation holds between the amount of ice cream sold in a month, x , and the number of deaths due to drowning, y . Of course, no one should suppose this relation represents ice cream sales as a cause of drowning.

Another reason why the Single Factor Model is not supported by the regression analogy is that there is another interpretation of evolutionary theory—the Statistical Interpretation—that also subscribes to the regression analogy. It too holds that there is a functional relation—a non-causal relation of statistical dependence—between population change, y , and the amount of variation in fitness in a population, x . Any particular change in a population, y_i , can be conceptually decomposed into the value predicted by the fitness distribution, ax_i , and error, ε_i . The first term, ax_i , is selection; the second, ε_i , is drift. In this interpretation, however, the probabilistic relation between fitness distribution

Not A Sure Thing

(i.e. selection) and population change is a non-causal relation of statistical correlation. As regression relations, themselves, do not distinguish causal dependence from mere statistical dependence, the regression analogy is neutral between the Single Factor Model and the Statistical Interpretation.

The Single Factor Model and the Statistical Interpretation differ primarily in their degrees of causal commitment. The regression analogy at least serves to identify the crux of the dispute between them—*viz.* whether to read the functional relation between fitness distribution and population change causally.

However, the regression analogy also raises doubts that this debate can ever be settled. The reason is that given a regression relation

“...[o]ne needs, among other things, extra-statistical assumptions about which are potentially causally relevant variables if the regression analysis is going to be used to support causal claims. (Woodward 1988: 260).⁹

In order to draw the correct causal inferences from a regression, we must know something about the causal structure of the phenomenon under question. Unfortunately, in our case, there is no further causal information to appeal to. The debate about the interpretation of fitness is part of a wider debate about which of the variables used to describe population change are “causally relevant”. Specifically, causal and statistical interpretations differ over whether *any* of the relevant population-level parameters are causes. The Statistical Interpretation holds that the only causes of population-level change are the causes of individual births, deaths and reproductions; the population-level

⁹ These “extra-statistical” assumptions, Woodward also calls “causal assumptions” (1988: 259).

Not A Sure Thing

parameters by which we describe population change, are not causal parameters. The causal interpretation holds that in addition to the individual-level causes there are population-level causes of selection and/or drift; at least the population parameters by which we describe population change *are* “causally relevant variables” (Brandon and Ramsey 2007, Reisman and Forber 2005, Shapiro and Sober 2007, Haug 2007).¹⁰ If we have to know the causes of evolution in order to know whether to read the regression analogy in the way recommended by the Single Factor Model or by the Statistical Interpretation, then there seems to be no non-question begging way to settle the dispute.

However, I believe the debate between the Single Factor Model and the Statistical Interpretation can be settled without importing these “extra-statistical” assumptions, again, by appealing to the Gillespie conception of fitness. An argument can be made that interpreting the distribution of Gillespie fitnesses as a cause of population change is incoherent; it leads to a particularly virulent form of Simpson’s paradox. Whereas most Simpson’s paradoxes can be resolved to yield a coherent causal story, this one cannot. I shall preface my argument a quick excursus through Simpson’s paradox.

¹⁰ Rosenberg (2006) and Bouchard and Rosenberg (2004) occupy an interesting intermediate position. They appear to agree with the statistical interpretation that all the causation goes on at the individual level. But, unlike both interpretations, they believe that trait fitness plays no unreduced explanatory role in evolutionary theory. The basic explanatory concept is individual (“ecological”) fitness.

4 Simpson's Paradox

Simpson's paradox is one of the potential pitfalls of inferring causes from probabilistic relations.¹¹ It occurs when for some putative cause C and its effect E ,

$$(4) P(E|C) > P(E|\sim C)$$

yet, for some exhaustive division of the population into sub-populations, F_1, \dots, F_n , for each sub-population, F_i ,

$$(5) P(E|C, F_i) < P(E|\sim C, F_i).$$

The reversal of probabilistic inequalities in (4) and (5) is known as a 'Simpson's reversal'. Simpson's reversal is a benign, workaday probabilistic phenomenon. It only causes difficulties when we attempt to draw causal inferences from probabilistic inequalities.

It is easy to illustrate the discomfiture that Simpson's reversal, like the one found in (4) and (5), can introduce into causal reasoning. Consider the Paradox of the Perplexing Painkiller. A series of drug trials suggest that, in the population overall, the probability of recovering from a headache, E , is raised by treatment with some new analgaesic drug, C , (as in 4). The results also show that when the test sample is divided up according to sex, for both men and women, the probability of non-treated patients recovering is higher than the probability of treated patients recovering, as per (5). A physician attempting to use these results in her clinical practice would encounter some peculiar problems. If a patient comes into her clinic complaining of the ailment and the physician does not know the sex of the patient, then she should treat her patient as a representative of the population as a whole, in which case the results suggest that she

¹¹ The discussion in this section draws heavily on Pearl's (2000), particularly Chapter 7.

Not A Sure Thing

should administering the drug, C (by 4). At the same time, she knows that the patient is either male or female, and if the patient is male, she shouldn't administer the drug (by 5) and if the patient is female she shouldn't administer the drug either (by 5). So, on the one hand, what the physician doesn't know about her patient changes her view on the effectiveness of the drug. At the same time she knows (from 5) that what she doesn't know is irrelevant to the effectiveness of the drug. Something has gone wrong. Our physician has an incoherent set of causal beliefs. She is embroiled in a Simpson's paradox.

Pearl (2000) offers an example illustrating how this sort of Simpson's reversal can arise and how the paradox can be resolved. Table 1 gives some hypothetical results from the problematic drug trials adapted from his discussion.

$F(\text{male})$	E	$\sim E$	n	Recovery
C	24	16	40	62%
$\sim C$	8	2	10	80%
$\sim F(\text{female})$	E	$\sim E$	n	Recovery
C	1	9	10	10%
$\sim C$	10	30	40	25%
Overall	E	$\sim E$	n	Recovery
C	20	20	50	50%
$\sim C$	18	32	50	36%

Table 1. Experiment 1: The effects of analgaesic C on headaches E

There is clearly a reversal of probabilistic inequalities here, as can be seen by comparing the recovery rates in F , $\sim F$, and Overall.

One of the causes of the reversal is the inequality among the sample sizes between treatments. In Experiment 1, treated males (40) comprise 40% of the sample; untreated

Not A Sure Thing

females comprise a further 40%. Sample bias seems to be a prevalent problem in meta-analysis of drug trial statistics (Hanley et al 2000). A study can inoculate itself against an unwanted Simpson's reversal by ensuring that sample sizes are constant between treatments (Hanley et al 2000).

Still, this prophylactic measure isn't failsafe and it isn't always available. In cases where the reversal of probabilistic inequalities does occur, we need a procedure for deciding which probabilities can be interpreted as causal and which cannot. Pearl (2000) tells us that we can use auxiliary information about the causal structure of the experiment as a guide. For example, we know a few causal facts about sex and drugs, like that taking an analgaesic typically doesn't cause one's sex. So, being a member of F or $\sim F$ is causally independent of the treatment, C . But, being male or female can have consequences for the probability of undergoing treatment. In this experiment, males are more likely than females to undergo treatment ($P(C|F) = .64 > P(C|\sim F) = .20$). Sex can also have consequences for the likelihood of recovery. Table 1 shows that males are more likely to recover than females whether or not they take the drug ($P(E|F) = .60 > P(E|\sim F) = .22$). The property that distinguishes the subpopulations has independent consequences both for the probability of C and the $P(E|C)$. Pearl depicts the causal structure of this experiment in the following way.

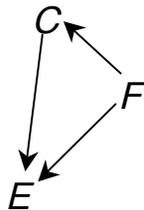


Figure 1. The causal relations between treatment with analgaesic, C , sex, F , and recovery from headache, E .

Not A Sure Thing

Given these considerations, Pearl suggests, we should consider the overall result, $P(E|C) = .50 > P(E|\sim C) = .36$, a statistical artefact. Table 1 does not accurately represent the causal relation between C and E in the population overall. It obscures the independent effects of F on both $P(C)$ and $P(E)$. Overall recovery rates are skewed by the fact that 80% of those who took the drug probably would have recovered whether they had or not and 80% of those who didn't take the drug probably wouldn't have recovered either way. We should infer that this drug does not relieve the affliction in the population overall—*contra* the suggestion in Table 1.

The moral to be drawn from this story is *not* that whenever there is a Simpson's reversal the overall effect is an artefact. Nor is it that wherever there is a Simpson's reversal *some* of the probabilistic relations must be non-causal. A minor change to the above example demonstrates this. Consider the *unparadoxical* case of the Ungentle Unguent. We are testing the efficacy of a skin cream (C) for the treatment of eczema (E). We find that some of the treated subjects develop a fever (F), more than we would expect. The results are given in Table 2.

F (fever)	E	$\sim E$	n	Recovery
C	24	16	40	62%
$\sim C$	8	2	10	80%
$\sim F$ (no fever)	E	$\sim E$	n	Recovery
C	1	9	10	10%
$\sim C$	10	30	40	25%
Overall	E	$\sim E$	n	Recovery
C	20	20	50	50%
$\sim C$	18	32	50	36%

Table 2. Experiment 2: the effect of skin cream, C , on eczema, E . Note that the values in the cells are identical to those of Table 2.

Not A Sure Thing

Note that in this experiment, the results are exactly as they were in the analgaesic experiment. The difference between Table 1 and Table 2 is not to be found in the values in the cells. The difference is “extra-statistical”.

In contrast to Experiment 1, we should not conclude from this experiment that the overall result is a statistical artefact. The reason is that in Experiment 2, unlike Experiment 1, the relation between F (fever) and C (the cream) is plausibly interpreted as a causal relation. In each experiment, the relation is expressed as

$$(6) \quad P(F|C)=.8 > P(F|\sim C)=.2$$

But whereas in Experiment 1, (6) does not express a causal relation, in Experiment 2 it does. The analgaesic, C , doesn't cause the patient's sex (Experiment 1). It is at least plausible to suppose that administering the cream, C , causes fever, F (Experiment 2). In Experiment 2, the treatment, C , changes the distribution of the subpopulations, F . It causes some subjects to be in subpopulation F who wouldn't otherwise be there. The treatment causes fever and fever independently raises the chances of recovery, E . The causal structure should be depicted as follows:

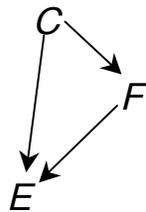


Figure 2. The causal relation between skin cream, C , relief from eczema, E , and fever, F , in Experiment 2.

Not A Sure Thing

The physician faced with these test results should administer the drug. She knows that it works, but through the ungentle means of inducing a fever.¹²

This scenario is similar to one discussed by Nancy Cartwright (1979) in which smoking (C) raises the chances of heart disease ($\sim E$) in those who exercise (F) and those who don't ($\sim F$). Yet, smoking causes people to exercise and, exercise prevents heart disease. It is reasonable in these cases to interpret *both* the within group effect and the overall effect as expressive of causal relations between C and E . Simpson's reversal occurs in these cases, but there is nothing paradoxical about interpreting all the conditional probabilities as representing causal relations.

The difference between the pathological cases of Simpson's reversal (Experiment 1) and the benign ones (Experiment 2) is to be found in their causal structures. In Experiment 1, F is causally independent of C ; in Experiment 2, it is not. But this information is not encoded in the conditional probabilities. One lesson to be learned is that chance raising alone is neither necessary nor sufficient for causation.

Causes can increase the probability of their effects; but they need not. And for the other way around: an increase in probability can be due to a causal connection; but lots of other things can be responsible as well (Cartwright 2001: 271).

Another lesson is that Simpson's paradox arises from the indiscriminate interpretation of conditional probabilities as causal relations. How to be more discriminating in our causal inferences is another question altogether.

¹² Notice here that, unlike in our first example, knowing the sub-population that the patient belongs to does make a difference to whether the physician should administer the treatment. If the patient has a fever already, the cream will lower the chances of recovery.

4.2 *The Sure Thing Principle*

It would help to know why the Simpson's paradox cases are paradoxical and why we so easily succumb to them. Pearl (2000) attributes our susceptibility to the paradoxes to a generalized human—perhaps better, a 'Humean'—psychological proclivity to interpret the probabilities in the 'calculus of proportions' by default as probabilities in the 'calculus of causes':

...humans are generally oblivious to rates and proportions ... and ... constantly search for causal relations Once people interpret proportions as causal relations, they continue to process those relations by causal calculus and not by the calculus of proportions. ... Were our minds governed by the calculus of proportions, ... Simpson's paradox would never have generated the attention that it did. (Pearl 2000:181)

The principal difference between the 'calculus of causes' and the 'calculus of proportions' is that the reversal of probabilistic inequalities is consistent with the calculus of proportions and sometimes *inconsistent* with the calculus of causes. This is because the calculus of causes conforms to the Sure Thing Principle (Pearl 2000). The Sure Thing Principle was originally formulated as precept of rational choice theory (Savage 1954).

If you would definitely prefer g to f , either knowing that the event C obtained, or knowing that event C did not obtain, then you definitely prefer g to f (Savage 1954: 21-22)

Not A Sure Thing

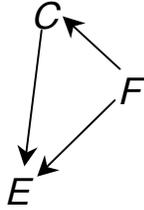
If you prefer maple walnut to vanilla either when you know it's summer or when you know it isn't summer, then you definitely prefer maple walnut to vanilla. Given the choice you should opt for maple walnut, whether or not you know what season it is.

Pearl formulates the causal version of the Sure Thing Principal in the following way:

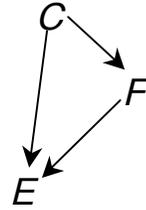
STP: An action C that increases the probability of event E in each subpopulation increases the probability of E in the population as a whole, provided that the action does not change the distribution of the subpopulations. (Pearl 2000:181).

The paradoxical cases are the ones that violate the causal version of the Sure Thing Principle. When causal inferences violate STP, we have an incoherent set of causal beliefs. This is readily apparent in the Paradox of the Perplexing Painkiller, for example. Here, an action, C (administering the drug), that increases the probability of *non*-recovery, $\sim E$, in each subpopulation, decreases it overall. Interpreting the probabilities in the Ungentle Unguent example as causal relations, however, is consistent with STP. STP does not apply to this system, because the proviso is not met—administering the treatment, C , changes the distribution of the subpopulations, F , (by raising the chances of fever). A quick comparison of Figures 1 and 2 confirms the difference:

Not A Sure Thing



Experiment 1 F is causally independent of C . The action C does not affect the distribution of the subpopulations, F and $\sim F$.



Experiment 2. F is causally dependent on C . The action C changes the distribution of the subpopulations, F and $\sim F$.

STP is the causal acid test. Every causal interpretation of the conditional probabilities that is inconsistent with STP is to be rejected. In each of the cases above we have used the auxiliary information about the causal structure of the experiment to determine whether there is a causal interpretation of the probabilities that is consistent with STP. In each case we found one. It is conceivable, however, that for some Simpson's reversals there is no plausible causal interpretation of the probabilities that is consistent with STP. In such a circumstance, there would be no coherent causal interpretation to be had. I believe that this kind of scenario can be constructed for the causal interpretation of fitness.

5. Simpson Meets Gillespie

We saw that where there is within generation variation in reproductive output, the Gillespie conception of fitness

$$(1) \quad w_i = \mu_i - \sigma_i^2/n$$

Not A Sure Thing

entails that variance and population size influence fitness. Immediately upon giving his measure of fitness, Gillespie cites a counterintuitive consequence.

With this definition comical situations will sometimes arise. For example, if $\mu_1 > \mu_2$ and $\sigma_1^2 < \sigma_2^2$, there exists a population size for which the two alleles are neutral. (Gillespie 1974: 604)

The following model illustrates one of Gillespie's 'comical situations'. Let the distribution of reproductive outputs be as follows:

Trait 1: $\mu_1=4.9, \sigma_1^2=0.7$

Trait 2: $\mu_2=5.0, \sigma_2^2=1.5$

The fitnesses of Trait's 1 and 2 can be plotted against population size as follows:

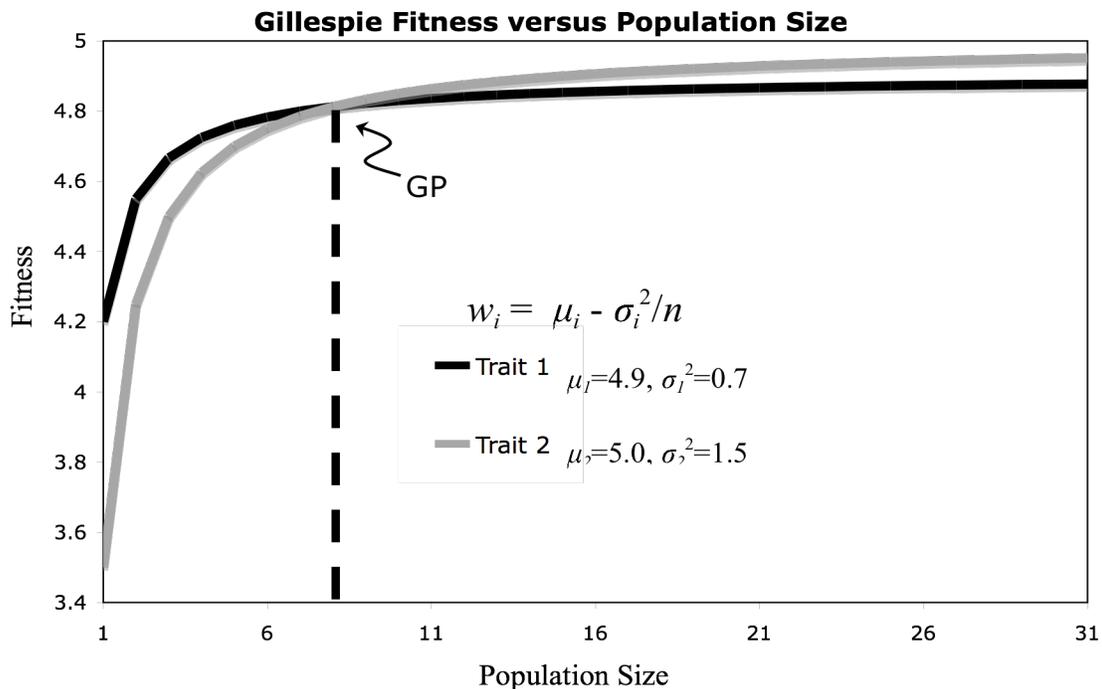


Figure 4. Gillespie fitness as a function of population size

Not A Sure Thing

GP is the ‘Gillespie Point’, the point at which $w_1=w_2$ ($n=8$). Where a single population has a Gillespie Point a further significant consequence arises. There is a reversal of fitness inequalities. In this model, where $n<8$, Trait₁ is fitter than Trait₂; where $n>8$ Trait₂ is fitter than Trait₁.

Suppose we have a homogeneous population, characterized as in Figure 4, comprising 14 subpopulations of 6 individuals each, in such a way that each has individuals of both Trait₁ and Trait₂. In each of these subpopulations, i , the fitness of Trait₁ exceeds that of Trait₂.

$$(7) \quad w_{1,i} > w_{2,i}$$

So long as the distributions of reproductive outputs (the μ_i s and σ_i^2 s) in the population overall are representative of the subpopulations, in each subpopulation, it will be more likely that Trait₁ increases in frequency relative to Trait₂ than vice versa. We are impelled to say that in each subpopulation there is selection for Trait₁ over Trait₂. But this is not merely an aggregate of 14 subpopulations, it is also a single population, o , of 64 organisms. In the population overall the fitness of Trait₂ exceeds that of Trait₁.

$$(8) \quad w_{2,o} > w_{1,o}$$

There is selection for Trait₂ over Trait₁.

Clearly, (7) and (8) constitute a Simpson’s reversal. Adopting our previous notation: let E be Trait₂ increases, C be the within subpopulation fitness distributions. Let $\sim C$ be some suitable null hypothesis— H_0 : $w_1=w_2$, and let the F_i s be the subpopulations.

$$(9) \quad P(E|C, F_i) < P(E|\sim C, F_i)$$

... ...

Not A Sure Thing

$$(10) \quad P(E|C, F_{14}) < P(E|\sim C, F_{14})$$

$$(11) \quad P(E|C) > P(E|\sim C)$$

This is no mere parochial, anomalous case. Every population can be represented as either a single large population or as an aggregate of smaller populations. In most populations, the magnitude of selection overall will differ from the sum of the magnitudes of selection within the subpopulations. The only special feature of this example is the reversal of fitness inequalities.

The causal interpretation of fitness enjoins us to read the probabilistic relation between fitness distribution and population change as causal. When the fitness of Trait₁ exceeds the fitness of Trait₂ (9, 10), there is an ensemble-level causal process—selection—that *causes* Trait₁ to preponderate over Trait₂. The causal interpretation of fitness, then, is committed to saying that within each subpopulation selection *causes* Trait₁ to increase over Trait₂ and that in the population overall (11), selection *causes* Trait₂ to increase over Trait₁. This looks like a Simpson’s paradox, a violation of the Sure Thing Principle. The causal interpretation appears to take on an incoherent set of causal assumptions. Something needs to be done.

The causal interpretation might try to explain away either the subpopulation probabilities or the overall probability as artefact, as was done in the Paradox of the Perplexing Painkiller. However, the usual remedies provide no relief here. For example, the reversal cannot be attributed to skewed sample sizes. The subpopulations are all the same size. Thus there is no ‘illegitimate averaging’ across unequal treatments.¹³ Nor can it be argued that there is a differential effect of subpopulation membership, F , on the

¹³ I take the expression “illegitimate averaging” from Glymour (1999)

Not A Sure Thing

value of C as there was in the Perplexing Painkiller case (see Table 1 and Fig. 1). In that instance, the reversal is attributable to the fact that the property that distinguishes F (being male) raises the prevalence of both C and E , whereas the property that distinguishes $\sim F$ (being female) lowers both C and E . But our biological population is disanalogous for three reasons. First, there is no difference in the value of the putative causal parameter, C , between subpopulations. So the reversal of probabilistic inequalities cannot be attributed to some independent causal factor that *differentially* affects C in each treatment. Second, the only plausible property of subpopulations, F , that could influence fitness distribution, C , is subpopulation size. Subpopulation size does have an influence on C , but this relation is constitutive, not causal. As we saw in section 3, intervening on population size doesn't cause a change in fitness distribution; it just *is* a change in fitness distribution. Third, it is incoherent to suppose that the effect of subpopulation size on fitness distribution is causal, on pain of incurring another violation of STP. If F (dividing the population into equal subpopulations) raises the chances of C in every subpopulation equally and C raises the chances of E in each subpopulation, then F raises the chances of E in each subpopulation. So F raises the chances of Trait₁ increasing over Trait₂ in each subpopulation (9, 10), but lowers chances of Trait₁ increasing over Trait₂ overall (11). This is yet another Simpson's paradox.

This is *disanalogous* to Experiment 1, in which F (being male) increases recovery rate in *all* sub-populations (both treated, C and non-treated, $\sim C$) *and* raises the recovery rate overall. There is no Simpson's reversal here. This is why it is coherent to interpret being male as a cause of recovery, but incoherent to interpret subpopulation membership

Not A Sure Thing

as a cause of population change. The upshot is that one cannot salvage the causal interpretation by explaining away the reversal of fitness inequalities as statistical artefact.

More importantly, however, we shouldn't *want* to explain away the reversal of fitness inequalities. There is nothing spurious about either the subpopulation fitness distribution or the overall population distribution. In order to get an accurate picture of the dynamics of the population we need both. For any subpopulation of $n < 8$, Trait₁ really is more likely to increase relative to Trait₂. In the population overall, Trait₂ really is more likely to increase relative to Trait₁.¹⁴ There is nothing pathological about the reversal of probabilistic inequalities. These really are the fitness distributions and they should not be explained away. The causal interpretation should want to retain the idea that *both* the subpopulation fitnesses and the overall fitnesses represent the causal efficacy of selection.

The only strategy remaining to the causal interpretation for relieving the Simpson's paradox is to invoke the proviso stated in STP: *viz.* 'provided that the action *does not change* the distribution of the subpopulations'. (Pearl 2000:181 emphasis added).¹⁵ If the proviso is not met, the Sure Thing Principle does not apply. In that case interpreting the reversal of probabilistic inequalities as a reversal of causal relations is consistent with STP. But, this defence of the causal interpretation of fitness is futile. The

¹⁴ This in itself sounds paradoxical. Michael Strevens (unpublished) gives an explanation. It is (roughly) that Trait₁ wins over Trait₂ in more subpopulations than Trait₂ wins over Trait₁. But when Trait₂ wins it tends to do so by a larger margin than when Trait₁ wins.

¹⁵ The proviso, recall, was the clause that licensed us to read both the within treatment probabilities and the overall probabilities as causal in the *unparadoxical* case of the Ungentle Unguent.

Not A Sure Thing

proviso is clearly met. Fitness distribution within the sub-populations, C , *does not change*, subpopulation size, F . The relation between fitness and population size is not analogous to the relation between applying the cream, C , and fever, F , in the Ungentle Unguent example, or the relation between smoking, C , and exercise, F , in Cartwright's (1979) scenario. In the biological example, C does not cause E *by* causing some intermediate effect F . The proviso offers no succour to the causal interpretation.

The upshot is that in our Gillespie experiment, interpreting the conditional probabilities in the calculus of causes commits us to a Simpson's paradox. There is no causal interpretation of the fitnesses that does justice to their explanatory role *and* is consistent with the Sure Thing Principle. Consequently, interpreting fitness distributions as causes leads to an incoherent set of causal commitments.

It is perfectly coherent, however, to interpret the probabilistic relation between fitness distribution in the calculus of proportions. Fitness distributions correlate with, but do not cause, the population changes they explain. Interpreting fitnesses as mere statistical correlations has the distinct advantage of allowing us to hold onto all the genuinely explanatory probability relations. It allows us to maintain consistently that within each subpopulation Trait₁ is likely to increase over Trait₂ (9, 10), but in the population overall Trait₂ is likely to increase over Trait₁ (11). Any other interpretation of the probabilities fails to do justice to the predictive and explanatory role of fitness.

This result should not only relieve us of any lingering inclination to interpret fitness, and the explanations it figures in, along the lines of the Two Factor Model, it puts paid to the Single Factor Model too. The feature of that model that distinguishes it from the Statistical Interpretation is that The Single Factor Model interprets the relation

Not A Sure Thing

between fitness distribution and population change as causal, whereas the Statistical interpretation takes it to be a mere statistical correlation. Given the flaws of the Two Factor and Single Factor models, the only candidate left standing is the Statistical Interpretation. It is the interpretation that remains after the excess causal commitments of, first, the Two Factor Model and then the Single Factor Model are stripped away. These are mere metaphysical excrescences, and should be removed from our interpretation of evolutionary theory.

Conclusion

Three competing interpretations of fitness, and the explanations it figures in, differ in their grades of causal commitment. The Two Factor Model holds that selection and drift are distinct causes of population change. The mechanism of selection is fitness distribution; the mechanism of selection is population size. The best method we have for individuating causes fails to demonstrate that drift and selection are distinct. Drift is not a separate cause of population change. The Single Factor Model takes on fewer causal commitments. It posits selection as a probabilistic cause of population change, but drift as merely error. The problem with this interpretation, one that also infects the Two Factor Model, is that fitness distribution fails to meet the requirements on being a cause. It violates the ‘calculus of causes’. Fitness is not a causal property of a trait type. Fitness distribution is not a cause of population change.

The only remaining alternative is the least causally committed interpretation—the Statistical Interpretation. Fitness distribution correlates with population change, but does not cause it. A natural selection explanation articulates a probabilistic relation between

Not A Sure Thing

fitness and population change. Natural selection theory predicts and explains population change without citing its causes.

The case against both of the causal interpretations turns on the Gillespie conception of fitness. Gillespie fitness demonstrates that the statistical parameter that best explains and predicts population change incorporates an interaction term between variance in reproductive output and population size. It is this interaction term that induces fitness distribution *not* to behave as a causal property should.¹⁶ If fitness is Gillespie fitness, then fitness is not causal.

¹⁶ See Strevens (unpublished) for an independent (and very elegant) argument to the same conclusion.

References Cited

- Abrams, Marshall (2007a) Fitness and Propensity's Annulment? *Biology and Philosophy* 22:115–130
- Abrams, Marshall (2007b) How Do Natural Selection and Random Drift Interact? *Philosophy of Science* (forthcoming).
- Ariew, A. and R.C. Lewontin (2004) The Confusions of Fitness. *British Journal for the Philosophy of Science* 55: 347–363.
- Beatty, J. (1984) Chance and Natural Selection. *Journal of Philosophy*. 51: 183-211
- Beatty, J (1992) Fitness *In*. Fox Keller, E and E.A. Lloyd 1992 *Key Words in Evolutionary Biology*. Cambridge, Ma.: Harvard University Press. pp.,115-119
- Beatty, J. and Finsen. S. (1989). Rethinking the propensity interpretation -- a peek inside Pandora's box. *In* What the Philosophy of Biology Is, *In* Ruse, Michael (ed.) pp. 17-30. Dordrecht: Kluwer Publishers.
- Bouchard, F. and Rosenberg, A. (2004) Fitness, Probability and the Principles of Natural Selection. *British Journal for the Philosophy of Science* 55: 693–712.
- Brandon, R. (2005) The Difference Between Selection and Drift: a reply to Millstein. *Biology and Philosophy* 20:153-170.
- Brandon, Robert (2006), The Principle of Drift: Biology's First Law. *Journal of Philosophy* 103: 319–336.
- Brandon, Robert, and Grant Ramsey (2007), What's Wrong with the Emergentist Statistical Interpretation of Natural Selection and Random Drift. *In* Michael Ruse

Not A Sure Thing

- and David L. Hull (eds.), *The Cambridge Companion to Philosophy of Biology*.
Cambridge: Cambridge University Press.
- Brandon, Robert, and John Beatty (1984), "The Propensity Interpretation of 'Fitness'—
No Interpretation Is No Substitute". *Philosophy of Science* 51: 342–347.
- Cartwright, N (1979) Causal Laws and Effective Strategies. *Nous* 13: 419-437.
- Dobzhansky, Theodosius and Olga Pavlovsky (1957), "An Experimental Study of the
Interaction Between Genetic Drift and Natural Selection." *Evolution* 11(3): 311-
19.
- Gillespie, J.H. (1974) Natural Selection for Within-Generation Variance in Offspring
Number. *Genetics* 76:601-606.
- Gillespie, J. H. (1973) Polymorphism in Random Environments. *Theoretical Population
Biology* 4:193-195.
- Gillespie, J. H. (1977). Natural selection for variances in offspring numbers: A new
evolutionary principle. *American Naturalist* 111, 1010–1014.
- Glymour, B. (1999) Population Level causation and a Unified Theory of Natural
Selection. *Biology and Philosophy* 14:521-536
- Grant, V. (1963) *The Origin of Adaptation*. New York: Columbia University Press.
- Hanley, J.A., Thériault, G., Reintjes, R. and de Boer A. (2000) Simpson's paradox in
Meta-Analysis. *Epidemiology*, Vol. 11, No. 5. (Sep., 2000), pp. 613-614.
- Haug, M. (2007) Of Mice and metaphysics: Natural Selection and Realized Population-
Level Properties. *Philosophy of Science*. 74: 431-451

Not A Sure Thing

Hodge, M.J.S. (1987) Natural Selection as a causal, Empirical and Probabilistic Theory.

In Kruger, L. (ed) 1987 *The Probabilistic Revolution*. Cambridge, Ma: MIT

Press. pp. 233-270.

Matthen, Mohan, and André Ariew (2002) Two Ways of Thinking about Fitness and

Natural Selection. *Journal of Philosophy* 99: 55–83

Millstein, R. L. (2002) Are Random Drift and Natural Selection Conceptually Distinct?

Biology and Philosophy 17:33–53.

Millstein, R. L. (2006) Natural Selection as a Population-Level causal process. *British*

Journal for the Philosophy of Science. 57:627-653.

Pearl J. (2000) *Causality*. Cambridge: Cambridge University Press.

Reisman, Kenneth, and Patrick Forber (2005) Manipulation and the Causes of Evolution.

Philosophy of Science 72: 1113–1123.

Rosenberg, Alexander (2006) *Darwinian Reductionism, or How to Stop Worrying and*

Love Molecular Biology. Chicago: University of Chicago Press.

Savage, L. J. (1954) *The Foundations of Statistics* New York: John Wiley and Sons.

Shapiro, Lawrence A. and Sober, E. (2007) Epiphenomenalism—The Do's and don'ts.

In G. Wolters and P. Machamer (eds), *Studies in Causality: Historical and*

Contemporary. University of Pittsburgh Press. 2007, pp. 235-264

Sober, Elliott (1984) *The Nature of Selection*. Cambridge, MA: MIT Press.

Not A Sure Thing

- Sober, E. (2001). The two faces of fitness. In R. S. Singh, C. B. Krimbas, D. B. Paul, and J. Beatty (Eds.), *Thinking About Evolution*, pp. 309–321. Cambridge University Press.
- Sober, E. (2008) *Evidence and Evolution: The Logic behind the Science*. Cambridge: Cambridge University Press.
- Stephens, Christopher (2004) Selection, Drift, and the ‘Forces’ of Evolution. *Philosophy of Science* 71: 550–570.
- Strevens, Michael, (unpublished) A Note on the Connection between Variance and Fitness” ms.
- Walsh, D.M. 2007 “The Pomp of superfluous Causes: The Interpretation of Evolutionary Theory” *Philosophy of Science* 74: 281-303
- Walsh, D. M., Tim Lewens, and André Ariew (2002) “The Trials of Life: Natural Selection and Random Drift”. *Philosophy of Science* 69: 452–473.
- Woodward, J. (1988) Understanding regression *PSA* 1: 255-269
- Woodward, J. (2002) What is a mechanism? A Counterfactual Account. *Philosophy of Science*, 69: S366–S377.
- Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Wright, S. (1931) Evolution in Mendelian populations. *Genetics* 16:97-159
- Wright, S. (1948) On the Roles of Directed and Random Changes in Gene Frequency in the Genetics of Populations *Evolution* 2:279-294.