

Selection in a Complex World: Deriving Causality from Stable Equilibrium

Hugh Desmond¹ 

Received: 19 April 2016 / Accepted: 21 February 2017 / Published online: 15 March 2017
© Springer Science+Business Media Dordrecht 2017

Abstract It is an ongoing controversy whether natural selection is a cause of population change, or a mere statistical description of how individual births and deaths accumulate. In this paper I restate the problem in terms of the reference class problem, and propose how the structure of stable equilibrium can provide a solution in continuity with biological practice. Insofar natural selection can be understood as a tendency towards equilibrium, key statisticalist criticisms are avoided. Further, in a modification of the Newtonian-force analogy, it can be suggested that a better metaphor for natural selection is that of an emergent force, similar in nature to entropic forces: with magnitude and direction, but lacking a spatiotemporal origin or point of application.

1 Introduction

Natural selection, both in biology textbooks and philosophical representations (e.g. Sober 1984), is often represented as some kind of Newtonian force, with magnitude and direction, originating in fitness differences and driving evolutionary change. By contrast, this metaphor has been rejected by the ‘statisticalist’ view of natural selection (e.g. Matthen and Ariew 2002, 2009; Walsh et al. 2002), which claims natural selection to be a mere book-keeping device, a way to keep count of the genuinely causal interactions that take place between individual organisms. Not even a cause, selection is an epiphenomenon, useful only for explanatory purposes.

While there are multiple ways of parsing this debate, the approach I will be adopting is to examine how the statisticalist view arises from the *complexity* of the organism-environment interaction and from ensuing reference class problems.

✉ Hugh Desmond
hugh.desmond@hiw.kuleuven.be

¹ Higher Institute of Philosophy, KULeuven, Blijde Inkomststraat 108/11, 3000 Leuven, Belgium

Complexity here refers to the very large number of independent processes that affect how often an organism will have reproduced by the end of its life, and the reference class problems that arise from this reflect the absence of any privileged way of abstracting away from this complexity.

Such themes have been noted before (for example, Brandon 1990, 2005; Kaplan 2013; Strevens 2016), but remain, on the whole, underemphasized. On the basis of a few assumptions about organism-environment interaction, a principled argument for the statisticalist view can be reconstructed, somewhat in the following vein (to be laid out in detail later): most of the organism-environment interactions that affect reproductive outcome are abstracted away from in an explanation by natural selection, but it is the sum total of such difference-making interactions that in fact causes populations to change. Hence natural selection is an epiphenomenon without any causal impact, a statistical apportioning of the fundamental causal reality.

Some of the main causalist counterarguments, even when the force analogy is abandoned, fail to make inroads on this ('no-privileged-abstraction') argument. One such counterargument is that natural selection is a cause insofar as intervening on natural selection at the population level 'makes a difference' in how populations evolve (Reisman and Forber 2005; Millstein 2006; Shapiro and Sober 2007; Glennan 2008; Gildenhuys 2014). Examples of interventions include modifying the selection coefficients and fitness values (Reisman and Forber 2007; Shapiro and Sober 2007), or introducing new phenotypes (Millstein 2006). This counterargument was rejected by Matthen and Ariew (2009) and Walsh (2007, 2010), who pointed out in various ways that the correlations between natural selection and different population parameters are mathematical in nature, not causal.

Furthermore, attributing fitness values to a group of individual organisms already presupposes a certain abstraction away from the complexity of organism-environment interaction, and thus such counterarguments beg the question. Hence, the question to be focused on instead is: are the abstractions involved in explanations by natural selection problematic for selection's supposedly causal nature?

I will seek to undermine the no-privileged-abstraction (NPA) argument by considering a class of counterexample: tendencies towards stable equilibrium. A population tending towards stable equilibrium is strong evidence for natural selection (Endler 1986); such tendencies also allow for certain abstractions to be privileged over others. In particular, the individual-level processes which do not make any causal difference at the longest time-scales (over multiple generations)—and thus do not make a difference to the outcome state—may be abstracted away from. In such cases natural selection may be understood as a force, with definite direction and magnitude, driving populations towards equilibrium.

Many if not most instances of evolution by natural selection do not actually approach stable equilibrium: environments may continually changing, and populations undergoing frequency-dependent selection may fail to settle down even in a constant environment. I will argue how the equilibrium model may be extended to such cases, so that the causal nature of selection in these cases may also understood as the tendency towards stable equilibrium, even though equilibrium is never realized.

An intuitive draw to the account presented here is that it implies an alternative (or modified¹) metaphor to that of Newtonian force. Natural selection is closer in nature to entropic forces, which are non-fundamental forces that originate in a system's statistical tendency to evolve towards a configuration with higher entropy.²

How precisely the concept of entropy can be applied to evolutionary process is an ongoing research programme (see e.g. Brooks and Wiley 1988; Barton and Coe 2009). What is important is that the analogy of entropic force captures two features of natural selection that the analogy with Newtonian forces does not: the lack of a localizable origin and the lack of a spatiotemporal point of application. Nonetheless, entropic forces have real consequences (e.g. osmosis, the Casimir effect) and thus are not simply epiphenomenal. In this way a key causalist intuition is supported.

The paper is structured as follows: in the first section I will make a principled case for the statisticalist position, and in the second section will begin to undermine it in light of how field biologists test the presence of natural selection. In the third and fourth section I present a model of natural selection as a causal tendency towards stable equilibrium, and deal with potential objections. The final section is a discussion of force metaphors, and of an extension of the model to frequency-dependent selection.

2 The No Privileged Abstraction Argument

The interaction between an organism and its environment is, generally speaking, complex. More often than not, environments are chaotic, characterized by ever-changing weather, geography and ecology. An organism itself is composed of many interacting parts, across numerous hierarchical structures—organ systems, organs, tissues, cells, nutrient gradients, etc. The resulting interaction between organism and environment has such a vast number of degrees of freedom as to make any parametrization practically impossible. Interactions can be biotic, such as competitive, symbiotic and parasitic interactions, or abiotic, such as extracting resources from the environment, or being affected by fundamental forces in the environment. While ecological models focus on only a few of these interactions at a time, in general we can assume that the number of degrees of freedom characterizing the actual organism-environment interaction is very large.³

However, not all processes will be causally relevant for the reproductive outcome of an organism. For example, consider a population of camouflaged and non-camouflaged moths. We can likely ignore how atmospheric pressure affects how much energy a moth expends when flapping its wings; we can likely ignore how slight variations in temperature have subtle effects on a moth's metabolic rate.

¹ There is no consensus as to what precisely are the defining conditions (e.g. additivity) of Newtonian forces. A detailed discussion of this would bring us too far from the main argument of this paper, but see Wilson (2007) or Stephens (2010).

² Verlinde (2011) describes an entropic force as an “effective macroscopic force that originates in a system with many degrees of freedom by the statistical tendency to increase its entropy.”

³ To emphasize: complexity is here taken to refer to the number of degrees of freedom, not functional complexity, or number of part-types (see Strevens 2003; McShea 2000).

Instead, the vicinity of a moth to a forest fire or to a predator are variables we will want to include when estimating a moth's likely reproductive outcome.

Yet even among these difference-making interactions between organism and environment, many will be abstracted away from in explanations by natural selection. The vicinity of a moth to a forest fire is likely to be ignored—unless, of course, the moth possesses some (perverse) heritable trait that makes it more likely to seek out fires. Once-off interactions are ignored, and interactions linked to the traits of an organism are included. On what basis are such traits selected, and what is their relation to the fitness (understood as the expected reproductive outcome⁴) of the organism?

The relation between trait and organism fitness is a difficult one—and, judging from a recent exchange between Sober (2013) and Pence and Ramsey (2013a), the two concepts seem intertwined. Pence and Ramsey have argued that definitions of trait fitness depend on organismic fitness, roughly because relevance for organismic reproductive success is the only measure by which to judge the relevance of a trait. By contrast, Sober has argued that organismic fitness is an actuarial quantity, estimated on the basis of how various traits affect organismic reproductive outcome.

Regardless of which fitness concept is the more fundamental one, statisticalists have denied that either measure of fitness can be estimated in a non-arbitrary way. Matthen (2009) has proposed that fitness measures are obtained by means of the condition of *metaconstancy*: a process may be excluded from consideration only if it is neutral with respect to (i.e. probabilistically uninfluenced by) all *heritable* properties. There is no heritable trait of a moth that correlates with its proximity to forest fires, and hence the evolutionary biologist may ignore it. Thus, in explanations by natural selection the causal difference-making processes that do not correlate with a heritable trait are to be ignored (see Fig. 1).

A fully detailed discussion of the implications of the reference class problem for fitness is beyond the scope of this paper. However, the following brief discussion of individual and trait fitness may convey the weight of the problem. First, the consequence for individual fitness would be that there is no privileged way of calculating the probability that an organism X will have n offspring (or have a daughter population of some size at $t = \infty$). This can be understood in the following way. If we exclude all non-metaconstant processes, we may arrive at a probability p ; however, if we take some other non-metaconstant process (e.g. the strength of the wind) into consideration, we might estimate a probability $p' \neq p$. Even if we ignore all non-metaconstant processes, there still be ambiguity as to the fitness value.

Could one not take *all* difference-making processes, both metaconstant and non-metaconstant, into consideration in the definition of individual fitness? This would be a strategy parallel to the well-known (but ultimately flawed) prescription of taking all evidence into account in determining the reference class (Carnap 1947) or taking the narrowest reference for which reliable statistics can still be compiled (Reichenbach 1949). Such a strategy would inherit the general problems of Carnap's or Reichenbach's strategies (see Hájek 2007); one problem in particular can be

⁴ There are more accurate and sophisticated measures of fitness available (such as Pence and Ramsey 2013b): see the discussion of fitness models later on.

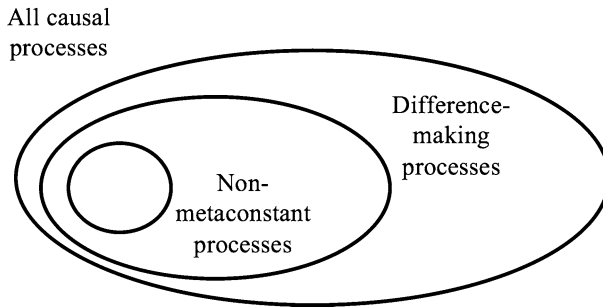


Fig. 1 Fitness abstracts away from most causal processes

pointed out. For example, consider individual A that is caught up in a forest fire and dies without offspring; individual A' with identical traits but a different spatial position avoids the forest fire and reproduces successfully. If one includes all difference-making processes in the definition of an individual fitness propensity, A and A' will be placed in different reference classes and will have different fitness values. In this way the difference between selection and drift evaporates and fitness ceases to be an explanatory notion.

Trait fitness faces similar issues. The probability of an individual having n offspring given trait T may depend on an indefinite number of other *unspecified* traits T_1, T_2, \dots , which may codify both metaconstant and non-metaconstant organism-environment interactions. Depending on which of the other traits T_1, T_2, \dots an organism possesses, a different probability will be obtained. If we simply include *all* possible traits (including ephemeral, non-heritable properties, such as the spatial position of a moth at a given moment in time), the distinction between selection for a trait and drift will collapse (see “[Appendix](#)”). Even restricting attention to all possible heritable traits, the resulting fitness value may not necessarily be explanatory of the dynamics in the actual population. For example, if fireproof, camouflaged moths have much smaller fitness than fireproof, non-camouflaged moths, this may lead to the trait of camouflage having a smaller fitness than the trait of non-camouflage in the hypothetical population (in contrast to the situation in the actual population, where fireproofness is absent).

Fundamentally, this argument is a restatement of the reference class problem in biological context. Individual organisms can be classified into groups in many different ways, due to the complexity of individual-environment interaction, and hence there is the problem how the probabilities underlying the actual births and deaths should be estimated. However, even if general resolutions to the reference class problem may be lacking (Hájek 2007), this does not entail that a more tailored response to the NPA argument should not be possible.

In order to deepen the connection between statisticalism and the NPA argument, I will finish the section by showing how three core statisticalist claims can be reconstructed as corollaries of the NPA argument:

(1) *Fitness can only be estimated through regression.* Since some difference-making processes are excluded from the explanation by abstraction, there is no a

priori way of estimating their impact on the probabilities of surviving and reproducing. Is the probability of having two offspring given the trait of camouflage .9 or .5? The value of that probability will be affected by the various causal processes impacting the individual, and these impacts will necessarily remain unknown since they were deliberately abstracted away from (Matthen 2009). Fitness values can be estimated only retrospectively, by a statistical regression on the actually occurred births and deaths in a population. Only insofar it may be used to extrapolate from past individual events can fitness have some predictive value. In this way, fitness values are mere statistical descriptors, not causes of actual frequencies of births and deaths.⁵

(2) *Fitness components are not additive.* The probabilistic relevance of having both traits T_1 and T_2 cannot be predicted from the relevance that T_1 and T_2 have by themselves, because the traits may be causally dependent (Matthen and Ariew 2002, 2009). For example (adapted from Stephens 2004), having resistance to malaria may increase fitness by p and being able to run fast may increase fitness by q , but in general we cannot know what fitness an organism will have when it has both traits, due to the unknown interaction between the two traits. Selection is not a causal force that can be simply divided into components.

(3) *Natural selection refers to the mathematical relation between change in frequencies and variance in rates, exemplified by Li's theorem.* Selection is variation in fitness, and since fitness is only a statistical descriptor of actual population, selection is not distinct from population change. To say a population is changing and to say individuals are reproducing at differential rates is to say the same thing. The upshot is that selection is not a cause mediating between fitness differences and population change in the way that the gravitational force mediates between mass distribution and acceleration of bodies (as represented in Sober 1984); rather it is more like the shadow of a flagpole (Walsh 2000; Matthen and Ariew 2009).

This is also why, in this view, causalist appeals to interventionist causality fail (Woodward 2003; Reisman and Forber 2005; Millstein 2006; Shapiro and Sober 2007; Glennan 2008): while it is conceivable to intervene on the gravitational force alone while keeping mass distribution constant (for example by changing the value of the gravitational constant), it is not conceivable to manipulate selection without changing fitness differences. Neither does the argument imply that there are no population-level causes (as claimed in Shapiro and Sober 2007): the epiphenomenality of selection is *not* a consequence of its population-level nature, but of its mathematical nature.

In summary, the NPA argument can be represented in the following format:

⁵ Note that in this way I do not view what is sometimes termed the 'metrological' and 'conceptual' roles of fitness (Pence and Ramsey 2013a) as wholly independent. Constructing quantitative fitness measures and estimating fitness values is surely a separate endeavor from investigating whether fitness as a concept allows for some causal interpretation. Nonetheless, the statisticalist argument against the *possibility* of a privileged estimation of fitness as quantitative measure—i.e., the argument that any quantitative measure of fitness is unavoidably dependent on explanatory interests—casts serious doubt on fitness as causal concept.

1. The space of life histories Ω is characterized a high number of degrees of freedom, where each independent variable makes a difference for the reproductive outcome of a life history.
2. Fitness (expected reproductive outcome) is specified by ignoring all degrees of freedom except a select few, dependent on explanatory interests.
3. Yet evolutionary change, being constituted by individual births and deaths, is affected by *all* degrees of freedom of Ω , not a subset of them (from (1)).
4. \therefore Natural selection, defined as variation in fitness, describes a particular pattern in population change, but since it does not refer to the processes that define Ω , it does not correspond to a cause of population change (from (2) and (3)).

3 Methodological Routes to Natural Selection

In light of the preceding argument, the interpretation of selection's causal nature can be reformulated in methodological terms: can selection (and drift) *in principle* be estimated from complex, natural populations in a way that is not hopelessly dependent on the explanatory interests a biologist might happen to have? In other words, is there a privileged way of statistically apportioning the underlying causal reality, or are there a plurality of ways, such that selection may assume any direction or magnitude as explanatory interests change?

While a general solution to the reference class problem for unconditional probabilities is likely impossible (Hájek 2007), I will argue there is a way out in the particular case of evolution by natural selection. In this section I will lay the ground by showing how previous key causalist arguments presuppose a successful resolution of the NPA argument and thus do not really address the NPA argument. Instead, it is instructive to examine how field biologists test for selection in natural populations. These field tests illustrate a class of case where the abstraction involved can be argued to be unambiguous and unproblematic.

3.1 Null Models, Selection Coefficients, Taxonomies and Fitness Models

One way the causality for selection has been argued for is by taking deviations from certain null models, such as Hardy–Weinberg equilibria, as sufficient evidence of the causal impact and force-like character of selection (e.g. Sober 1984; Stephens 2004). Even if one does not hold to the force analogy, it has also been argued that such deviation establishes that natural selection must be an objective cause of evolutionary change, and not merely some arbitrary statistical redescription of individual births and deaths (Gildenhuis 2014).

However, this line of argument does not help, because the question is whether such null models are the result of arbitrary abstraction. Hence they cannot be used to defend a causal interpretation of natural selection since the ambiguity in divvying up the individual-level causal processes remains. For instance, if the frequency of the recessive homozygous genotype *aa* is higher than the one expected by the Hardy–Weinberg model, this only means that the genotype *aa* is correlated with a

larger number of offspring in the population under consideration. It may turn out that some other allele B is not present at all in the population, and if it would be, that different correlations between aa and offspring number would be obtained (see discussion about trait fitness in previous section). Thus the test does not establish an objective, privileged fitness differential between aa and the other genotypes.

A similar problem arises in the causalist strategy of drawing on direct estimates of selection, through response to selection (by means of the breeder's equation) or selection coefficients. For instance, since manipulating selection coefficients correlates with population change, this manipulation has been taken as a sign that selection must be causal (e.g. Forber and Reisman 2007).

However, such a strategy merely shifts the problem from the interpretation of selection to the interpretation of the selection coefficients, and the NPA argument can be likewise applied to the selection coefficients. Coefficients are not manipulated directly: what is manipulated in practice are traits, trait frequencies and trait-environment interactions, and that such manipulations lead to population change is perfectly compatible with a statisticalist interpretation of the selection coefficients. Whether or not changes in selection coefficients can be estimated in a privileged way given such manipulations is precisely the point of contention.

Another strategy is to propose a taxonomy of fitness measures, and to claim differences in one type of fitness are causes of evolution (e.g. Sober 1984, 2013; Abrams 2012; Ramsey 2013). For instance, Abrams (2012) distinguishes between various kinds of fitness measure, and argues that 'parametric type fitness' may be given a causal interpretation, since this fitness parameter is what underlies the empirical estimations of trait fitness values in actual populations.

However, such a taxonomy does not prevent the NPA argument from being applied to the probabilities underlying parametric fitness (see discussion of trait fitness in introduction). Abrams implicitly acknowledges this point when mentioning that the degree of certainty of our knowledge of parametric type fitness is directly influenced by the method of estimation (Abrams 2012, 196: 4). However, a thoroughgoing response to the NPA argument must involve delving deeper into the details of these methodological inferences.

Fitness models, which integrate a relatively wide range of causal influences, face a similar problem. For instance, the Pence–Ramsey model of fitness (Pence and Ramsey 2013b) integrates a variety of causal processes that had been wrongfully ignored by the original propensity interpretation of fitness (Brandon 1978; Mills and Beatty 1979): processes that affect the variance and skew (and higher moments) in reproductive outcomes, processes that affect the expected reproductive outcome of offspring generations (but not the parent generation), and processes that affect the timing when reproduction happens within the life-history of a generation. These processes do not directly affect the expected number of offspring of the parent generation, but do affect the expected size of the daughter population many generations in the future.

However, while the Pence–Ramsey model is more inclusive of processes that affect the ways in which probabilities are distributed over reproductive outcomes and how this distribution changes over time, it does not address the origin of these probabilities themselves. In assigning probabilities to possible daughter populations (at infinite time), the skeptical NPA argument could once again be applied to show

how explanatory interests influence the estimation of such probabilities (see discussion of individual fitness in introduction).

3.2 Natural Populations

The approach that I suggest in this paper is that we should consider methods used by field biologists to establish the presence of selection. Such methods strongly suggest that, at least in certain cases, there is a non-arbitrary way out of the reference class problem. How precisely natural selection may be considered a cause in such cases, and how precisely this answers statisticalist objects are separate questions, and are the subjects of the next two sections.

In field studies, estimating selection coefficients is often very difficult because of the variability in the temporal dynamics of population change. Siepielski et al. (2009), in a review of replicated field studies, indicate that, in general, natural selection varies in strength, direction (positive and negative) and form (linear and nonlinear selection) from one generation to the next.

This variability in selection dynamics is a reflection of the complexity of individual-environment interaction. Environments are rarely static, and slight changes in biotic or abiotic variables may have large effects on the patterns of births and deaths exhibited. Ideally we should be able to exactly replicate population and environmental structure as to establish whether population changes are simply noise or follow an expected pattern; however, in the vast majority of cases this is not possible. Even when adequate temporal replicates exist, the duration of the study is very rarely longer than a couple of years (Siepielski et al. 2009).

In such cases, where variability is so ubiquitous, it becomes difficult to rebut statisticalist challenges. Many of the assumptions in population genetics models (infinite populations, static environments, absence of linkage, and so on) are made in order to gain mathematical tractability. However, while variability and complexity may be eliminated in abstract models, they are more or less permanent features of real environments, and one may legitimately question whether such abstract models can be used to judge on questions of causality when it is unclear whether even the basic inputs of the models (such as selection coefficients) reflect objective properties of organisms and populations.

Nonetheless, there is a powerful group of methods for cutting through short-term and individual-level complexity: tests for *stable equilibrium*. Endler (1986) describes two ways in which equilibria may be used to establish the presence of natural selection. The first method ('method V') requires longitudinal studies on trait frequency distributions, and tests either for long-term stability in frequencies, or steady directional change. Of the two, an observed directionality is taken as stronger evidence for natural selection. However, it is not fool-proof: Endler describes how simulations of coin-toss runs surprisingly often give rise to (temporary) directional trends. Even though a single unbiased coin-tossing experiment leads to heads as often as it does to tails, the probability of a significant run of either heads or tails is surprisingly high.

The second method ('method VI') has a manipulative character: a population at stasis is perturbed, and a subsequent observation of directional change, either back

to its previous state or towards a new state, is taken as strong evidence for natural selection. One implementation of the perturbation entails manipulating the trait frequencies of the population. Reversion to the previous distribution of frequencies is taken as strong evidence for the presence of natural selection. Changes in the environment constitute another type of perturbation. Such perturbations can be human induced (e.g. pesticides) or natural (e.g. volcanic eruptions, epidemics), and can occur abruptly or gradually. An example of where a gradual change in environment is used to establish selection is the well-known study of moth evolution in response to increasing air pollution in Britain (Kettlewell 1955).

Such methods are not always possible. Sometimes there are multiple equilibria in the system, complicating the interpretation. Further, not all forms of natural selection, such as certain types of frequency dependent selection, seem to tend towards equilibrium. Nonetheless, when there *is* a tendency towards equilibrium, it can be discerned in longitudinal studies as a long-term effective change in population structure, despite short-term fluctuations.

4 The Equilibrium Model of Causality

While such methods undercut the generality of the no-privileged-abstraction argument, in themselves they do not constitute a direct argument for the causality of selection. In this section such a direct argument will be attempted. We will assume that a natural population is undergoing *directional change* in its trait frequency distribution, and is evolving towards a stable equilibrium. This assumption may be thought of as the outcome of an empirical field investigation; in the next section the assumption will be discussed more critically, especially with regards to the worry that this introduces some arbitrary abstraction.

I will first argue that natural selection, as tendency towards equilibrium, is a cause according to the (probabilistic) counterfactual understanding of causality (e.g. Lewis 1973); subsequently I will argue it is a force with direction and magnitude.

4.1 Relation and Relata

Let us first briefly outline the landscape of accounts of the causality of selection with regard to two reference points: the causal relation and causal relata. This will allow for the view defended here to be situated with more precision.

Concerning relation, I subscribe to the statisticalist criticism that the relation is not one of causal *production* (Matthen and Ariew 2002): evolution by natural selection is not a spatiotemporally continuous process where some physical mark is transmitted, or quantity conserved (Salmon 1984; Dowe 2000). However, most causalist accounts have drawn on a difference-making relation. In light of problems arising from the use of the interventionalist criterion of causality to adjudicate on the epiphenomenality of population-level processes (Baumgartner 2010; Matthen and Ariew 2009), and in light of the fact that such interventions often depend on abstract models, I will avoid that specific conception of causality (as does e.g.

Huneman 2012). Instead, I will consider whether natural selection represents a probabilistic *counterfactual* relation between its relata (e.g. Lewis 1973).

Even within the counterfactual option there are a number of possible accounts according to how the relata—the source and effect of natural selection—are to be understood. One point of controversy concerns the *level of analysis* at which the source and effect of natural selection are to be located. Bouchard and Rosenberg (2004) propose that the source of natural selection consists in pair-wise individual-level competitive differences and that the effect consists in pair-wise differences in (expected) individual reproductive outcomes (Bouchard and Rosenberg 2004). I follow Millstein's (2006) criticism of this proposal, and consider natural selection at the population level.

A further point of contention is whether the source of selection is (population-level) fitness differences and whether the effect is changes in trait frequencies (e.g. Millstein 2006; Sober 1984). One problem here is that natural selection and changes in trait frequencies are mathematical, not causal, consequences of fitness differences (Matthen and Ariew 2009). Another problem, as argued previously in this paper, is that it is not clear whether fitness variables are objective in the same way mass or acceleration are, and hence to claim selection is causal because fitness values can be manipulated is to beg the question.

By contrast, the proposal of this paper can be summarized in following scheme:

population with trait distribution \mathbf{d} and in environment $E \rightarrow$ natural selection
 \rightarrow directional change in trait frequencies

The vector $\mathbf{d} = (d_1, d_2, \dots, d_n)$ specifies the relative frequency of each of the n traits that characterize organisms in the population. I will now argue that this proposal ensures that each of the relations (between selection and source, and between selection and effect) is causally counterfactual.

First, concerning the first relation, the combination of n traits and environment E gives rise to a great number of biotic and abiotic organism-environment interactions. Some of those interactions will make no net difference to the effective direction of population change, while others will. Natural selection is only constituted by the latter interactions.⁶

Understood in this way, natural selection is not a mathematical consequence of \mathbf{d} and E because the causal interactions between organisms of the population and between organisms and environment are not mathematical consequences of a specification of \mathbf{d} and E . (For example, if the laws of physics were changed, the same \mathbf{d} and E would give rise to different causal interactions.) Further, if \mathbf{d} and E were different, we would likely see different causal interactions, and thus a different process of natural selection.

Second, the effect of natural selection is to be sought in the *direction* of change, not the *fact* of change. This proposal contrasts with the accounts presented by Huneman (2012), Millstein (2006) and Matthen and Ariew (2009), who all locate

⁶ This is continuous with Huneman's (2012) definition of selection pressures as ecological "reliable factors which differentially affect the trait types" (185). The only difference is that reliableness is specifically defined here in reference to an effective directionality.

the effect of natural selection in the *fact* of population change. The problem with the latter proposal is that population change is just as likely to occur where natural selection is absent—such as when drift is present, or when the population evolves towards a Hardy–Weinberg equilibrium state. Hence the question to be asked is not whether selection causes an accumulation of births and deaths, but whether selection causes a stable trend in the accumulation of births and deaths.

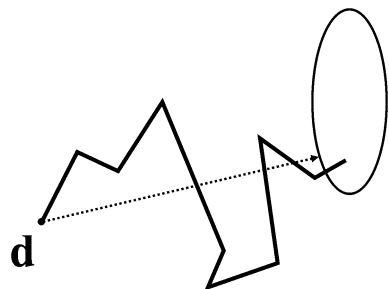
One could compare the effect of natural selection on a population to the effect of an electric field on the random walk of a charged particle in a gas. The gas has a certain temperature, and the particles collide randomly; however, because of the electric field there is a bias in a certain direction. The field does not cause the movement of the particle itself, but does make a difference for the long-term ‘effective’ path the particle will take (Fig. 2). In this way natural selection’s causal influence is to be sought in the effective direction of population change instead of change itself, as the latter could also be the result of random fluctuation (drift).

A special case occurs when the effective path shrinks to a point, that is to say, if the trait distribution \mathbf{d} of the population is already in a stable equilibrium state. This occurs with stabilizing selection, and in this case the causal effect of natural selection is to hold the population in stable equilibrium. This effect becomes observable when the trait distribution is perturbed to $\mathbf{d} + \delta\mathbf{d}$: given stabilizing selection, one would expect the population to evolve back towards equilibrium.

In general, the relation between natural selection and effect is causally counterfactual: were natural selection not present, directional evolution would be much less likely. Drift could in principle give rise to directional evolution, but the probability of this occurring diminishes the longer the directional change lasts. Further, natural selection is not sufficient for a directional change in frequencies taking hold, as natural selection can be counteracted by drift.

At this point, one may rehearse an objection, going back to Sober (1983), namely that equilibrium explanations are not causal (see also Huneman 2010 for a generalization of this argument). Sober argued that explanations where the outcome state is explained as a stable equilibrium state are not causal because such explanations do not pick out the *actual cause* of the outcome state. For example, if a marble is let go at the rim of a bowl, and proceeds to find its way to the lowest point of the bowl, we explain its outcome state by the shape of the bowl, not by the particular path the marble followed. In other words, we explain by referring to a disjunction of possible causal scenarios (1983: 84): if a system had not followed this

Fig. 2 Actual versus effective evolutionary change



particular pathway to equilibrium, it would have attained equilibrium by another path.

However, this does not present a problem, because Sober is drawing on a different notion of causality. If we mean by cause the actual, specific path a population takes on way to equilibrium, then indeed an explanation by natural selection does not pick out any actual cause. However, a counterfactual difference-making relation exists between the equilibrium state and the directionality of the *trend* leading up to the equilibrium. Explanations by natural selection are causal insofar they pick out directional trends as causes of the outcome state (equilibrium).

A second objection concerns how a population may undergo directional evolution towards equilibrium even without natural selection or any causal force acting on it. This is the case for the two main models of zero-force evolution: Hardy–Weinberg (H–W) equilibria (Sober 1984) and drift (Brandon 2006, McShea and Brandon 2010). In these models, the population either jumps to an equilibrium state (a H–W equilibrium), or tends towards a state where a single allele has become fixated in the population. This would seem to undermine the counterfactual relation between natural selection and stable equilibrium, since a population may spontaneously tend towards equilibrium in absence of natural selection or any causal force.

In response one can point out that neither zero-force model describes a stable equilibrium. In the case of drift, the absorbing state of fixation is not a stable equilibrium, but a neutral equilibrium state. Should the population be perturbed from the state of fixation, it would not necessarily drift back to that same state, but may drift to some other fixation state. Thus the process of drift is like a drunk man’s walk on a sidewalk with a gutter on either side. Once the drunk man falls into the gutter he stays there, but does not necessarily return to the gutter if pulled out by someone. There is no tendency towards falling into one particular gutter.

Similarly, a H–W equilibrium state is not stable either. If (p, q, r) is a H–W equilibrium state (p and r represent the relative frequencies of the homozygous traits AA and BB; q represents the relative frequency of the heterozygous trait AB), and is perturbed to $(p + \delta_p, q + \delta_q, r + \delta_r)$, then this perturbed state will only evolve back towards (p, q, r) if $\delta_p = \delta_r$ and $\delta_q = -2\delta_p$ (see “Appendix”). H–W equilibrium points are not stable equilibrium points, as they are stable only along one specific line in state space.

In this way, if natural selection were not present, regardless of which zero-force model one adheres to, there would not be a trend towards stable equilibrium. Drift may lead to population change, but only selection leads to a robust, multigenerational trend in the change of population structure.

4.2 Direction and Magnitude

In instances where natural selection causes a population to trend towards a stable equilibrium state, the stable equilibrium is a reference point that allows a direction to be ascribed to the population change. Note that, stabilizing, disrupting and directional selection all are ‘directional’ in this sense, since they tend either towards the fixation of a particular trait, or towards stable polymorphism.

The magnitude of selection can be estimated in the regular way, through the response to selection. The only difference in the equilibrium framework is that what matters is the *effective* response to selection. In the case only a single trait is selected, the breeder's equation is applicable.⁷

$$R_{eff} = H^2 s_{eff}$$

where H^2 is the heritability, R_{eff} the effective response to selection, and s_{eff} the (effective) strength of natural selection.

Thus, if a population never reaches the equilibrium state ($N \rightarrow \infty$), the effective response to selection is zero, even though the magnitude of selection may be nonzero. In this case the tendency towards the equilibrium state does not translate into any actual trend. This is a direct consequence of the breeder's equation, and some underlying reasons for this can be seen by separating heritability into genetic and environmental variation: $H^2 = V_G/(V_G + V_E)$ where V_G is the portion of total phenotypic variance that can be explained by genetic variance, and V_E is the portion explainable by variation in the environment. Thus, as drift increases, the environment will have an increasingly variable effect on the population, and hence the effective response to selection goes to zero. Likewise, if traits are transmitted poorly between generations, the response to selection will be diminished.

5 Statistalst Objections

In this section we consider two fundamental objections to the model. The first is that the model simply reintroduces a different arbitrary abstraction away from the complexity of organism-environment interaction. The second is that natural selection still has an epiphenomenal character.

5.1 Directionality and Abstraction

Even limiting attention to cases where populations trend to stable equilibrium, one objection could be that defining natural selection with respect to the effective direction of population change is simply a different way of arbitrarily abstracting away from the complexity of individual-level causal processes. Based on an observed approach to stable equilibrium, the equilibrium model distinguishes between processes that make a difference to an individual's reproductive outcome (Δ_2) and those that make a difference to the stable equilibrium (Δ_3 —see Fig. 3). However, or so the objection goes, in principle some explanatory interest in some *other* feature of the population change could lead to some different distinction being

⁷ Determining the response to selection when traits are correlated through nonadditive genetic effects is more complicated, but there are methods using multivariate regression, such as Lande's equation: $R = \mathbf{G}\mathbf{P}^{-1}\mathbf{s}$, where \mathbf{G} and \mathbf{P} are the additive genetic and phenotypic variance-covariance matrices. In its canonical formulation, this equation describes the response within a generation (see Lande and Arnold 1983), so it could be conceivably extended to describe the effective response over multiple generations by means of an effective selection coefficient vector: $R_{eff} = \mathbf{G}\mathbf{P}^{-1}\mathbf{s}_{eff}$.

made among causal processes (Δ'_2 and Δ'_3), and some other direction (if any at all) to be ascribed to the population change.

In response, one must first emphasize that the issue is not whether abstraction can be avoided in explanations by natural selection; it is whether the abstraction involved precludes a causal interpretation of explanations by natural selection. Hence the real problem that must be considered is whether some arbitrary explanatory interests are smuggled into the equilibrium model.

To further specify the locus of the objection, note that not all dependence on explanatory interests is automatically problematic. When observing the behavior of natural populations, an evolutionary biologist will be interested in explaining different types of outcome than, say, a biochemist, or an insurance agent. Explanatory interests in such cases dictate the type of outcome that is considered the explanandum. What is at stake is that, even when the outcome of interest is decided on, whether explanatory interests may lead to the set of Δ_3 being defined differently, so that there is no privileged way of (causally) explaining the evolutionary change.

The equilibrium model avoids this problem, since Δ_3 is not obtained from Δ_2 by arbitrary selection, but by *averaging out* across many generations. Those causal processes which do not average out to zero are part of Δ_3 , and these are the processes which make no long-term difference to the structure of the population. Explanatory interests cannot alter what makes a difference to the equilibrium state. It is an objective matter of fact which processes do and which processes do not make an overall difference. Furthermore, since equilibrium can be observed independently of probability estimations, observations of equilibrium can serve as a basis for a privileged estimation of the probabilities underlying fitness, allowing for a way out of the NPA argument in these cases.

In this respect, the equilibrium model follows the causal interpretation of the abstraction involved in explanations by natural selection (Strevens 2009; Woodward 2003): the individual-level interactions which average-out to zero are those which are not relevant difference-makers, and hence may be abstracted away from. Even though abstraction may lead to simplified or even distorted representations of causal processes, this is justified because the abstraction cuts the causal difference-making structure 'at the joints'.

To what extent idealized models of population-level trends in general may be considered causal is a question beyond the scope of the paper (but see Levins 1966; Weisberg 2007). For instance various models in ecology do not represent individual-level interactions veridically, such as Hubbell's neutral theory (Hubbell 2001), which treats all individuals in a community as equivalent. Yet, the resulting model is deemed surprisingly predictive of population-level biodiversity patterns (Chave 2004). Likewise, models in population genetics take only a limited number of loci into consideration, often (though not always) assume constant environments (see Gillespie 2004), and yet have been used very successfully to understand the evolution of actual populations.

Given that idealized models typically cannot explain every aspect of a complex phenomenon (so that some such as Levins 1966 have advocated the use of a cluster of models), it is unsurprising that some explanatory-dependence cannot be avoided,

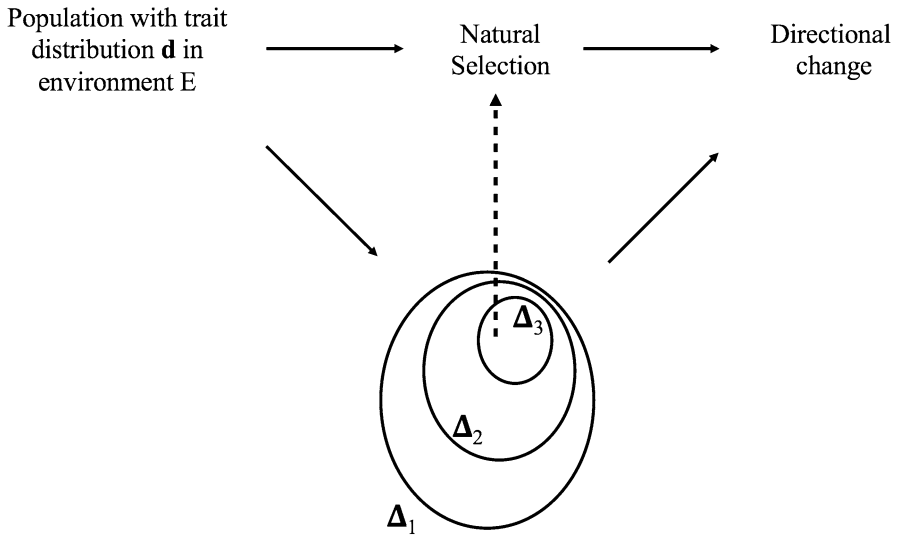


Fig. 3 The *solid lines* represent causal relations; the *dashed line* represents a constitutive relation. Δ_1 represents the set of all individual-level causal processes (defining a specific life-history), Δ_2 the set of all individual-level processes affecting the reproductive outcome of a life-history, and Δ_3 the processes which make a difference to the equilibrium distribution of trait frequencies

since one type of idealization is chosen over alternative types. We mentioned the choice of outcome state, and the other important dependence on explanatory interests allowed for by the equilibrium model is that on *time-scale*. What makes a difference on the long-term does not necessarily make a relevant difference on the short-term, and vice versa (see next section).⁸ However, once the target and scope of the explanation is set, the equilibrium model does not possess any of the arbitrary description-dependence assumed by the NPA argument.

5.2 Epiphenomenalism

One of key statisticalist challenges was to point to the epiphenomenal character of natural selection: selection is a *tertium quid* that could be eliminated from consideration without affecting the causes actually driving evolution. The original version of this challenge focused on the mathematical character of natural selection, and while we already argued why natural selection is not a mathematical consequence of the source of selection, one could still object that the causal relations connecting organism-environment structure, natural selection and directional change could be eliminated without affecting directional change.

Figure 3 schematically represents this objection: if natural selection were to be eliminated from consideration, the causal relation between the processes determining individual births and deaths, Δ_2 , and directional change would still be intact.

⁸ See also Abrams (2009) for a discussion of how what Abrams calls the ‘reference environment’ is dependent on the target explanandum.

Evolution by natural selection is still ‘just’ an accumulation of births and deaths in the equilibrium model.

In response, one need only to point out that only the processes in Δ_3 , not Δ_2 , cause *directional* change. Aspects of the causal processes impinging on organisms in a population may be changed without changing the long-term outcome, but the latter depends counterfactually on the long-term difference-making processes (Δ_3). To eliminate natural selection from the causal scheme is to eliminate the difference between Δ_2 and Δ_3 —and this affects the causal scheme.⁹

A related way of putting the same point is that natural selection contains *more* information about the causal structure of population change than contained by the mass of causal processes Δ_2 . Unlike Δ_2 , natural selection tells us something about *possible* evolutionary changes, not only about actual changes. Specifically, it tells that if the population would be perturbed, the population would tend to return to equilibrium, in virtue of the processes in Δ_3 . Eliminating natural selection would eliminate this information about the causal structure of population change.

6 Discussion: General Implications

6.1 What About Frequency-Dependence?

Throughout the paper we have alluded to cases such as frequency-dependent selection, where evolution by natural selection most often does not tend towards stable equilibrium (and even leads to ergodicity and chaotic dynamics: see for example Doebeli and Ispolatov 2014). In fact, many real-life cases do not involve any discernible approach towards stable equilibrium.

However, the equilibrium model may be extended to such cases by first distinguishing between the realization of equilibrium and tendency towards equilibrium. We already mentioned how the tendency towards equilibrium may be counteracted by drift, so that equilibrium is never attained, or so that a population may not even exhibit a discernible tendency towards equilibrium. However, in cases of frequency-dependence equilibrium may not be attained even if there is only selection acting on a population.

Hence a further distinction must be made between a variable component (variable direction) and an effective component (fixed direction) of natural selection. The effective component would describe approaches to stable equilibria, and in general, attractor states, while the variable component would describe the rest. If there is no convergence to stable equilibrium, the effective component is zero. This raises the question whether the equilibrium model allows for selection to be considered causal in such cases.

To reduce cases where selection has no privileged direction (i.e., the direction of selection averages out to zero over time) to the equilibrium model, one can analyze

⁹ In this respect, the account presented here is continuous with the suggestion that natural selection is a structuring cause in the space of population life-histories (Ramsey 2015). See also Kaplan (2013) for the suggestion that natural selection is some measure of robustness.

a single process of frequency-dependent selection as a succession of different selection processes, each tending towards (but not realizing) a stable equilibrium. For example, consider frequency-dependent selection for color in populations of *Poeciliid* fish (cf. Huneman 2012): predators are tuned to prey on fish with majority colors, so that any one color will undergo cycles of selection-for and selection-against, depending on whether it is a majority or minority color. Is the same instance of natural selection doing the selecting for and the selecting against? According to this extension of the equilibrium model, the answer is no, as the interaction between a fish with a given color and its environment changes in the two cases. Any cycle of selection-for and selection-against is a succession of selection processes each tending towards a different stable equilibrium.

To better understand this reduction, consider Brandon's distinction between 'homogeneous' and 'heterogeneous' selective environments (Brandon 2005). Fitness can only be ascribed to an organism or trait if the selective environment is 'homogeneous' within the relevant region. If the selective environment is heterogeneous, then the nature of the organism-environment interaction fluctuates significantly, and Brandon terms the process of selection that arises from such an environment 'compound selection'. Similarly, frequency-dependent selection could be understood as compound selection, where the selective environment is temporally heterogeneous (due to changing biotic interactions).

It may not be easy to distinguish between heterogeneous environments and fluctuating homogeneous environments. For example, a volcano eruption may be interpreted as a fluctuation on a longer time-scale, or as a stable but short-lived part of a heterogeneous succession of environments. In each case the population may be driven to extinction, but in the first case the process will be classified as drift and in the latter it may be classified as selection. A full discussion of this problem would go beyond the scope of this paper (but see e.g. Abrams 2009); more important is to emphasize that this ambiguity does not mean that the choice between the two interpretations depends on arbitrary explanatory interests. As Brandon (2005:166) notes, distinguishing between these cases will depend on the details of the population, such as behavioral patterns. In this way, an equilibrium may be stable on one scale and neutral on another: this means that the choice of appropriate scale should depend on the precise nature of the organism-environment interaction.

To summarize, variable selection may be analyzed as a succession of different instantaneous selection processes, each directed towards a different stable equilibrium that is never reached because the environment is continually changing. In this way, even though determining the precise nature of the selective environment may be difficult in practise, variable selection may be considered causal insofar it is a type of compound selection, where each instance of selection is covered by the equilibrium model.

6.2 What Kind of Force is Selection?

Finally, in order to both broaden the perspectives of the paper and to bring it to an intuitive close, I would wish to consider natural selection at a more metaphorical level. I have argued natural selection is a causal force on basis of certain formal

characteristics (counterfactual dependence, direction, magnitude); however, it is fair to wonder what kind of a force it is precisely. It is clearly not a classical Newtonian force: neither its source, nor its effect is spatiotemporally localizable. There is also no ‘field’ of force associated with natural selection. Is it possible to say something more positive about what kind of force it is?

In their 2002 paper, Matthen and Ariew compare evolution by natural selection to heat flow. During heat flow there is no transmission of a physical mark; it is instead a statistical process where the concentration of fast-moving molecules in the warm part of the gas spreads throughout the container. While the net result is that energy smoothly diffuses, in reality this occurs one discontinuous collision at a time. Similarly, evolution by natural selection is not the smooth diffusion of an advantageous trait, but is mediated by individual births and deaths.

From this perspective, the concept of entropic force has some suggestive power. Entropic forces are forces that originate in the statistical tendency of complex systems with many degrees of freedom to increase their entropy. Natural populations are not in thermodynamic equilibrium, and hence the concept of thermodynamic entropy has little meaning in this context. However, the analogy does hold insofar natural selection is an effective population-level force that originates in a tendency of the population to reorganize its structure through interaction with the environment.

Does this suggestion not undo the effort to argue against the epiphenomenalism of selection? Not necessarily, because even though entropic forces emerge out of lower-level interactions and thus are not fundamental, they can do work in displacing objects. Osmosis is an example of a process driven by an entropic force. Entropic forces have real effects, can be manipulated, and can be ascribed a direction and magnitude. Thus they meet the criteria of many accounts of difference-making, whether Woodward’s manipulationism or the probabilistic counterfactual account.

Thus the fact that an entropic force is not fundamental but emerges out of lower-level interactions does not in itself constitute an argument that such forces are epiphenomenal and non-causal. Interestingly, it may be pointed out that a strict distinction between a force being ‘fundamental’ and a ‘statistical byproduct’ is increasingly under pressure in contemporary physics. Numerous proposals have been made of how fundamental forces, including gravity, may actually be entropic in nature (Verlinde 2011; Freund 2010).

7 Conclusion

The complexity of the organism-environment interaction means that the resulting evolutionary change can, in principle, be statistically apportioned in many ways. While statisticalists have argued that this multiplicity implies that any single apportioning does not ‘cut nature at the joints’ and is instead dependent on arbitrary explanatory interests, I have argued that, at least in some cases, a privileged apportioning can be made. In cases where the population tends towards

stable equilibrium, a clear distinction can be made between individual-level interactions that do and do not make a difference to the outcome equilibrium state.

Cases where there is no discernible approach to equilibrium may be judged either as cases where selection is counteracted, or (such as in cases of frequency-dependent selection) as compound selection. The upshot is a vision of natural selection as an entropic force: a non-fundamental force, arising out of statistical tendencies at the level of individual interactions, with magnitude and direction.

Acknowledgments I would wish to thank Michael Strevens, Grant Ramsey, Andreas De Block, Charles Pence, Philippe Huneman and an anonymous referee for helpful comments on previous versions of this paper.

Appendix: Instability of H–W Equilibria

First, let us investigate when two distributions (p, q, r) and (p', q', r') will give rise to the same H–W equilibrium. Then the following three identities must hold:

$$\begin{aligned} p + q + r &= p' + q' + r' = 1 \\ p + q/2 &= p' + q'/2 \\ r + q/2 &= r' + q'/2 \end{aligned}$$

These equations are dependent, and taking r' as a parameter, we get the following set of solutions $\{(p - r + r', 1 - p + r - 2r', r') | r' \in [0, 1]\}$. This can be simplified with the change of variable $\delta = r' - r$, and thus we can say that the basin of the H–W equilibrium $((p + q/2)^2, 2(p + q/2)(r + q/2), (r + q/2)^2)$ is

$$\{(p + \delta, q - 2\delta, r + \delta) | \delta \in [-r, 1 - r]\}.$$

The basin of a single H–W equilibrium point is the line with direction $(1, -2, 1)$ in distribution space. As one would expect, by letting $\delta = (p + q/2)^2 - p$ one can see that the H–W equilibrium itself is part of its own basin.

Hence we may conclude that there is no open neighbourhood \mathcal{N} around any H–W point (p, q, r) such that \mathcal{N} is enclosed by the basin of (p, q, r) . H–W equilibria are stable only along one specific line, and hence are unstable equilibria.

References

- Abrams, M. (2009). What determines biological fitness? The problem of the reference environment. *Synthese*, 166, 21–40. doi:10.1007/s11229-007-9255-9.
- Abrams, M. (2012). Measured, modeled, and causal conceptions of fitness. *Frontiers in Genetics*, 3, 196. doi:10.3389/fgene.2012.00196.
- Barton, N. H., & Coe, J. B. (2009). On the application of statistical physics to evolutionary biology. *Journal of Theoretical Biology*. doi:10.1016/j.jtbi.2009.03.019.
- Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of Philosophy*, 40, 359–384.

- Bouchard, F., & Rosenberg, A. (2004). Fitness, probability and the principles of natural selection. *British Journal for Philosophy of Science*, 4, 693–712.
- Brandon, R. N. (1978). Adaptation and evolutionary theory. *Studies in History and Philosophy of Science Part A*, 9, 181–206.
- Brandon, R. N. (1990). *Adaptation and environment*. Princeton University Press.
- Brandon, R. N. (2005). The difference between selection and drift: A reply to millstein. *Biology and Philosophy*, 20, 153–170.
- Brandon, R. N. (2006). The principle of drift: Biology's first law. *Journal of Philosophy*, 7, 319–355.
- Brooks, D., & Wiley, E. (1988). *Evolution as entropy*. Science and its conceptual foundations series. Chicago: University of Chicago Press.
- Carnap, R. (1947). On the application of inductive logic. *Philosophy and Phenomenological Research*, 8, 133–148.
- Chave, J. (2004). Neutral theory and community ecology. *Ecology Letters*, 7, 241–253. doi:10.1111/j.1461-0248.2003.00566.x.
- Doebeli, M., & Ispolatov, I. (2014). Chaos and unpredictability in evolution. *Evolution*, 68(5), 1365–1373. doi:10.1111/evo.12354.
- Dowe, P. (2000). *Physical causation*. Cambridge: Cambridge University Press.
- Endler, J. A. (1986). *Natural selection in the wild*. Princeton: Princeton University Press.
- Forber, P., & K. Reisman (2007). Can there be stochastic evolutionary causes? *Philosophy of Science*, 74(5), 616–627.
- Freund, P. G. O. (2010). Emergent gauge fields. arXiv:1008.4147
- Gildenhuys, P. (2014). Arbitrariness and causation in classical population genetics. *British Journal for the Philosophy of Science*, 65, 429–444.
- Gillespie, J. H. (2004). *Population genetics: A concise guide*. Baltimore, MD: The Johns Hopkins University Press.
- Glennan, S. (2008). Productivity, relevance and natural selection. *Biology and Philosophy*, 24, 325–339.
- Hájek, A. (2007). The reference class problem is your problem too. *Synthese*, 156, 563–585.
- Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography*. Princeton: Princeton University Press.
- Huneman, Philippe. (2010). Topological explanations and robustness in biological sciences. *Synthese*, 177, 213–245.
- Huneman, Philippe. (2012). Natural selection: A case for the counterfactual approach. *Erkenntnis*, 76, 171–194.
- Kaplan, J. M. (2013). “Relevant similarity” and the causes of biological evolution: Selection, fitness, and statistically abstractive explanations. *Biology and Philosophy*, 28, 405–421. doi:10.1007/s10539-012-9342-2.
- Kettlewell, H. B. D. (1955). Selection experiments on industrial melanism in the lepidoptera. *Heredity*, 9, 323–342.
- Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54, 421–431.
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70, 556–567. (Reprinted with Postscripts in Lewis, D. (1986). *Philosophical Papers Volume II*, pp. 159–213, Oxford: Oxford University Press).
- Matthen, M. (2009). Drift and “statistically abstractive explanation”. *Philosophy of Science*, 76, 464–487.
- Matthen, M., & Ariew, A. (2002). Two ways of thinking about fitness and natural selection. *Journal of Philosophy*, 99, 55–83.
- Matthen, M., & Ariew, A. (2009). Selection and causation. *Philosophy of Science*, 76, 201–224.
- McShea, D. W. (2000). Functional complexity in organisms: Parts as proxies. *Biology and Philosophy*, 15(5), 641–668.
- McShea, D. W., & Brandon, R. N. (2010). *Biology's first law: The tendency for diversity and complexity to increase in evolutionary systems*. University of Chicago Press.
- Mills, S. A., & Beatty, J. H. (1979). The propensity interpretation of fitness. *Philosophy of Science*, 46, 263–286.
- Millstein, R. (2006). Natural selection as a population-level causal process. *British Journal for the Philosophy of Science*, 57(4), 627–653.
- Pence, C. H., & Ramsey, G. (2013a). Is organismic fitness at the basis of evolutionary theory? *Philosophy of Science*, 82, 1081–1091.
- Pence, C. H., & Ramsey, G. (2013b). A new foundation for the propensity interpretation of fitness. *The British Journal for the Philosophy of Science*, 64, 851–881.

- Ramsey, G. (2013). Can fitness differences be a cause of evolution? *Philosophy and Theory in Biology*, 5, e401.
- Ramsey, G. (2015). The causal structure of evolutionary theory. *Australasian Journal of Philosophy*. doi:10.1080/00048402.2015.1111398.
- Reichenbach, H. (1949). *The theory of probability*. Oakland: University of California Press.
- Reisman, K., & Forber, P. (2005). Manipulation and the causes of evolution. *Philosophy of Science*, 72, 1113–1123.
- Reisman, K., & Forber, P. (2007). Can there be stochastic evolutionary causes? *Philosophy of Science*, 74, 616–627.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.
- Shapiro, L., & Sober, E. (2007). Epiphenomenalism: The Do's and the Don'ts. In P. Machamer & G. Wolters (Eds.), *Thinking about causes: From Greek philosophy to modern physics* (pp. 235–264). Pittsburgh: University of Pittsburgh Press.
- Siepielski, A. M., DiBattista, J. D., & Carlson, S. M. (2009). It's about time: The temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, 12, 1261–1276. doi:10.1111/j.1461-0248.2009.01381.x.
- Sober, E. (1983). Equilibrium explanation. *Philosophical Studies*, 43(2), 201–210.
- Sober, E. (1984). *The nature of selection: Evolutionary theory in philosophical focus*. Cambridge, MA: Bradford Books, MIT Press.
- Sober, E. (2013). Trait fitness is not a propensity, but fitness variation is. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44, 336–341.
- Stephens, C. (2004). Selection, drift, and the “forces” of evolution. *Philosophy of Science*, 71, 550–570.
- Stephens, C. (2010). Forces and causes in evolutionary theory. *Philosophy of Science*, 77, 716–727.
- Strevens, M. (2003). *Bigger than chaos: Understanding complexity through probability*. Cambridge, MA: Harvard University Press.
- Strevens, M. (2009). *Depth: An account of scientific explanation*. Cambridge, MA: Harvard University Press.
- Strevens, M. (2016). The reference class problem in evolutionary biology: Distinguishing selection from drift. In C. Pence & G. Ramsey (Eds.), *Chance in evolution*. Chicago: University of Chicago Press.
- Verlinde, E. (2011). On the origin of gravity and the laws of Newton. *Journal for High Energy Physics*. doi:10.1007/JHEP04(2011)029.
- Walsh, D. (2000). Chasing shadows: Natural selection and adaptation. *Studies in the History and Philosophy of Biological and Biomedical Sciences*, 31, 135–153.
- Walsh, D. (2007). The pomp of superfluous causes: The interpretation of evolutionary theory. *Philosophy of Science*, 74, 281–303.
- Walsh, D. (2010). Not a sure thing: Fitness, probability and causation. *Philosophy of Science*, 77, 147–171.
- Walsh, D., Lewens, T., & Ariew, A. (2002). Trials of life: Natural selection and random drift. *Philosophy of Science*, 69, 452–473.
- Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, 104, 639–659.
- Wilson, J. (2007). Newtonian forces. *British Journal for Philosophy of Science*, 58, 173–205.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.