

# Fitness, Probability and the Principles of Natural Selection

Frédéric Bouchard and Alex Rosenberg

---

## ABSTRACT

We argue that a fashionable interpretation of the theory of natural selection as a claim exclusively about populations is mistaken. The interpretation rests on adopting an analysis of fitness as a probabilistic propensity which cannot be substantiated, draws parallels with thermodynamics which are without foundations, and fails to do justice to the fundamental distinction between drift and selection. This distinction requires a notion of fitness as a pairwise comparison between individuals taken two at a time, and so vitiates the interpretation of the theory as one about populations exclusively.

---

## Introduction

There is a strong temptation to treat the theory of natural selection solely as a claim about the ‘central tendencies’ in evolution. In the words of Sterelny and Kitcher ([1988], p. 345): ‘evolutionary theory, like statistical mechanics, has no use for such a fine grain of description [as the biography of each organism]: the aim is to make clear the central tendencies in the history of evolving populations.’ Not only does such an interpretation *seem* to do justice to the centrality of population genetics—a thoroughly statistical enterprise—to evolutionary biology, but it also bids fair to solve or dissolve the long-standing problem of explaining or explaining away ‘fitness’, and to account for the theory’s character in terms familiar from physics. In this paper we show why the temptation must be resisted. We argue that these benefits are not obtained, and that, moreover, the approach obscures crucial facts about the theory of natural selection. In Section 1 we show that the probabilistic propensity account of fitness required by the ‘central tendencies’ approach is no substitute for the causal conception of comparative fitness as a pairwise relation between individual organisms. In Section 2 we refute recent attempts to purge the theory of a causal concept of fitness by interpreting the theory, along the lines of the second law of thermodynamics, as one exclusively about ‘ensembles’.

In Section 3 we apply these results to an analysis of the nature of evolutionary drift, which both demonstrates the evolutionary indispensability of the causal concept of fitness and identifies the real source of evolutionary probabilities.

- 1 *Problems of probabilistic fitness*
  - 2 *The PNS and the second law of thermodynamics*
  - 3 *Drift and selection*
  - 4 *Conclusion*
- 

## 1 Problems of probabilistic fitness

It is convenient to express the ‘central tendency’ approach in terms of a contrast between two principles of natural selection.<sup>1</sup> One version quantifies over populations:

PNS<sub>pop</sub> (x) (y) (E) [If x and y are competing populations and x is fitter than y in E at generation n, then probably (x’s size is larger than y in E at some generation n’ later than n)]

The PNS<sub>pop</sub> is relativized to environments E, since fitness is relative to an environment.<sup>2</sup> We cannot narrow down the later generation beyond ‘some’ generation or other, for reasons that will be made clear below.

The ‘central tendencies’ interpretation of the theory of natural selection is the claim that the PNS<sub>pop</sub> need not be further grounded on claims about the fitness of individuals.<sup>3</sup> Biologists and philosophers who reject the temptation to stop with this principle may wish to endorse something like the following principle of natural selection for individuals:

<sup>1</sup> Cf. Brandon ([1978], [1990]) for discussion of these formulations of the Principle of Natural Selection, hereafter PNS.

<sup>2</sup> Some recent advocates of the ‘central tendencies’ approach substitute one of the PNS’s deductive consequences for the PNS itself. Instead of the PNS, Matthen and Ariew ([2002], p. 73) treat Fisher’s fundamental theorem of natural selection as the central explanatory principle of Darwinism. This theorem states that the fitness of a population increases at a rate proportional to the genetic variance in fitness present in the population (cf. Strickberger [1985], p. 728). Matthen and Ariew’s version of the theorem, is attributed to C .C. Li: ‘In a subdivided population the rate of change in [overall population] growth rate [i.e. fitness] is proportional to the variance in growth rates [i.e. fitnesses].’ ‘Variance’, of course, is a population-level property, and so suits Fisher’s theorem to express the central tendency thesis that Matthen and Ariew defend. However, the claims to be made here about the PNS<sub>pop</sub> could equally well be made with respect to Fisher’s theorem.

<sup>3</sup> Cf. also Walsh, Lewens and Ariew ([2002], p. 460): ‘The objective of natural selection is to explain and predict changes in the relative frequencies of heritable traits within a population. The change that selection explains is a consequence of *variation in* fitness (citing Lewontin [1970, 1974], Brandon [1990]).’ And further, Walsh, Lewens and Ariew ([2002], p. 469): ‘natural selection theory explains changes in the structure of a population, but not by appeal to the individual-level causes of births, deaths, and reproductions.’

$PNS_{ind}(x)(y)(E)$  [ $x$  and  $y$  are competing organisms in generation  $n$ , and  $x$  is fitter than  $y$  in  $E$ , then probably (there is some generation  $n'$ , at which  $x$  has more descendants than  $y$ )]

These biologists and philosophers hold that, when  $PNS_{pop}$  obtains, the  $PNS_{ind}$  is the most important part of the explanation as to why it obtains.<sup>4</sup>

Both principles employ the relational property ‘ $x$  is fitter than  $y$ ’ in their antecedents and the sentential operator ‘probably ( $p$ )’ in their consequents. What these terms mean remain two of the most vexed questions of the philosophy of biology. On the answers to them turn the status of the theory of natural selection as empirical science, and its connection to every other compartment of biology.

Here are two proposed answers to the question of what ‘fitness’ in the antecedent and ‘probably’ in the consequent of the  $PNS_{pop}$  (and the  $PNS_{ind}$  if there is one) mean that are popular among philosophers of biology:

$x$  is fitter than  $y$  in  $E$  iff  $x$  has a probabilistic propensity to leave more offspring in  $E$  greater than  $y$ ’s probabilistic propensity to leave more offspring in  $E$ .<sup>5</sup>

‘Probably(\_\_\_\_)’ = ‘The long-run relative frequency of (\_\_\_\_) approaches 1’.

If we plug these two proposals into either or both the  $PNS_{pop}$  and  $PNS_{ind}$ , at least three questions arise.

First and perhaps most obvious is the question of how the consequents of either  $PNS_{pop}$  or  $PNS_{ind}$  are related to the finite actual sequences if they are claims about the relative frequency in the long run, i.e. about infinite sequences. The problems here are well known.<sup>6</sup> We need a way of applying

<sup>4</sup> As Sober ([2002]) points out, fitness has two quite different roles in evolutionary theory: ‘[Fitness] describes the relationship of an organism to its environment. It also has a mathematical representation that allows predictions and explanations to be formulated’ (p. 309). The  $PNS_{pop}$  employs ‘the mathematical representation’. The  $PNS_{ind}$  employs the organism/environment relation which, it is worth noting, is closer to the way Darwin originally framed the theory.

<sup>5</sup> The *loci classici* of this definition are Brandon ([1978]), and Beatty and Mills ([1979]). See also Brandon ([1990], chapter 1), and Sober ([1993] and [2000], both p. 71). The propensity to leave more offspring than one’s competitor can also be understood as the summation of the propensities to leave specific numbers of offspring: if  $a$  has a propensity to leave 1 offspring higher than  $b$ ’s propensity to leave 1, and if  $a$  has a higher propensity to leave 2 offspring than  $b$ ’s propensity to leave 2 offspring, then we may summarize by saying that  $a$  has a higher propensity to leave more offspring than  $b$ ’s propensity to leave more offspring. It should be noted that such a propensity doesn’t have to focus on offspring numbers. Sober defines the comparative fitness of *traits*, not *individuals* or *populations*, stating ‘Trait X is fitter than trait Y if and only if X has a higher probability of survival and/or reproductive success than Y’ ([2000], p. 71). Traits are types, i.e. abstract properties. Their survival and/or expected reproductive success is a matter either of the tropes that instantiate them or of the individuals or populations which manifest these tropes and through them the traits. Trait fitness differences require individual or population fitness differences. Other propensity accounts have also focused on persistence of a type instead of offspring numbers exclusively (see for instance Thoday [1953]). In this paper, we use ‘propensity accounts’ as shorthand for propensity accounts focusing on offspring contribution.

<sup>6</sup> For a general introduction to these problems, see W. Salmon ([1966], pp. 83–95).

a claim about infinite sequences to the actual finite sequences which the PNSs are to explain. Although no non-controversial solution to this problem is available, there must be one. That the probability in the consequent is a propensity has already, so to speak, been spoken for. It is the interpretation of probability which almost invariably figures in the fitness-relations reported in antecedents of the  $\text{PNS}_{\text{pop}}$  and  $\text{PNS}_{\text{ind}}$ . If we adopt the same meaning for the probability in the consequent, then when the appropriate grammatical changes are made to accommodate this interpretation by attributing probabilistic dispositions to  $x$  and  $y$ , the two PNSs will turn out to be tautologies.

The second issue is closely related. For either version of the PNS to be a contingent truth there must be some factual *difference* between the two probabilities:

$x$  has a probabilistic propensity to leave more offspring than  $y$  greater than  $y$ 's probabilistic propensity to leave more offspring than  $x$  in every generation after  $n$

and

the long-run relative frequency of ( $x$ 's leaving more offspring than  $y$ ) approaches 1.

If there is no factual difference between these two probabilities, both versions of the PNS become tautologies. Another way to put the point is that, in the two PNSs, the antecedent is supposed to identify a cause and the consequent an effect. Accordingly there must be, at least in principle, a difference between them in conceptual if not empirical content. What would show that there is a difference between these two kinds of probabilities?

There certainly are philosophers of science who deny that an empirical distinction between probabilistic propensities and long-run relative frequencies is in general possible (see Earman [1986], pp. 147–51). Putting aside empiricist strictures, would it suffice to claim that, here as in quantum mechanics, we find a brute unanalyzable probabilistic dispositional property of a particular item, which generates long-run relative frequencies? Among philosophers of quantum mechanics, some hold that probabilistic propensities can explain actual frequencies (cf. Railton [1978], p. 216), and some hold that they do so via a detour into long-run relative frequencies. But, owing to empiricist commitments, few are comfortable with such arguments and adopt them only because, at the level of the quantum mechanical, probabilistic propensities are indispensable and irreducible (cf. Lewis [1984]). Proponents of probabilistic propensities in the PNSs may envision two possibilities here. One is that probabilistic propensities at the levels of phenomena that constitute the biological are the result of quantum probabilities 'percolating up' in Sober's ([1984]) and Brandon and Carson's ([1996]) phrase; the second is that there are brute unexplainable probabilistic propensities at the level of

organismal fitness differences. No one doubts the possibility of quantum percolation at the biological level. It may be a source of mutations (see Stamos [2001] for a discussion). But the claim that it has a *significant* role in fitness differences is not supported by any independent evidence (see Glymour [2001] for a discussion). The claim that there are brute probabilistic propensities at the level of organismal fitness differences (Brandon and Carson [1996]) is only slightly more tenable. No one has adduced any evidence that, for instance, the probabilistic generalizations about the behavior of animals which ethology and behavioral biology provide are irreducibly statistical. Rather, they are expressions of the current state of our knowledge and ignorance of the causes and conditions of the behavior in question. Empiricist-inspired suspicion of the explanatory power of dispositions that lack even possible manifest-trait foundations seems well grounded in biology.

These first two problems about probabilities in evolution are largely philosophical. The third issue facing any interpretation of the PNSs is a biologically urgent matter: it turns out to be difficult to pin down the specific probabilistic propensity that constitutes fitness altogether. The difficulty reflects features of natural selection that we must accommodate. And it leads inexorably to the conclusion that far from providing the theoretical meaning of fitness, the probabilistic propensity ‘definition’ is a set of an indefinitely large number of operational *measures* of fitness. Moreover, identifying which of these measures to use turns on prior determinations of whether natural selection obtains and what has been selected. The upshot will be that the probabilistic propensity ‘definition’ does not figure in either the  $PNS_{pop}$  or the  $PNS_{ind}$ .

The first thing to notice about the ‘definition’

‘ $x$  is fitter than  $y$  in  $E$ ’ = ‘ $x$  has a probabilistic propensity to leave more offspring than  $y$  in  $E$  > than  $y$ ’s probabilistic propensity to leave more offspring than  $x$  in  $E$ ’

is that it makes the PNSs into falsehoods. That is, there are many circumstances in which the organism with the higher number of expected offspring is the less fit organism, not the more fit. For example, Gillespie ([1977]) has shown there are cases in which the temporal and/or spatial variance in number of offspring may also have an important selective effect which swamps mere numbers in any given generation.<sup>7</sup> To accommodate these biological cases, we need to qualify the ‘definition’ to include the effects of variance.<sup>8</sup>

<sup>7</sup> To take a simple example from Brandon ([1990]), if organism  $a$  has 2 offspring each year, and organism  $b$  has 1 offspring in odd numbered years and 3 in even numbered ones, then, *ceteris paribus*, after ten generations there will be 512 descendants of  $a$  and 243 descendants of  $b$ . The same holds if  $a$  and  $b$  are populations, and  $b$ ’s offspring vary between 1 and 3 depending on location instead of period.

<sup>8</sup> Sober ([2002]) also notes that Gillespie introduces variance in order to show that fitness cannot be defined simply in terms of the expected number of offspring.

$x$  is fitter than  $y$  = probably  $x$  will have more offspring than  $y$ , unless their average numbers of offspring are equal and the temporal and/or spatial variance in  $y$ 's offspring numbers is greater than the variance in  $x$ 's, or the average numbers of  $x$ 's offspring are lower than  $y$ 's, but the difference in offspring variance is large enough to counterbalance  $y$ 's greater number of offspring.

It is also the case that in some biologically actual circumstances—for example, in circumstances in which mean fitnesses are low, increased variance is sometimes selected for (see Ekbohm, Fagerstrom and Agren [1980]). Indeed, as Beatty and Finsen ([1987]) have noted, sometimes the 'skew' or geometric means of offspring numbers and variance may effect selection. Thus, the 'definition' of fitness must take these conditions into account on pain of turning the PNSs into falsehoods. One simple way to protect the PNSs from falsehood is to add a *ceteris paribus* clause to the definition. But the question must then be raised of how many different exceptions to the original *definiens* need to be accommodated. If the circumstances under which greater offspring numbers do not make for greater fitness are indefinitely many, then this 'definition' will be unsatisfactory.

Some proponents of the propensity definition recognize this difficulty and are prepared to accept that at most a 'schematic' definition can be provided. Thus Brandon ([1990], p. 20) writes:

We can [...] define the adaptedness [a synonym for expected fitness] of an organism  $O$  in an environment  $E$  as follows:  $A^*(O,E) = \sigma P(Q_i^{OE}) - f(E, \sigma^2)$ .

Here  $Q_i^{OE}$  are a range of possible offspring numbers in generation  $i$ ,  $P(Q_i^{OE})$  is the probabilistic propensity to leave  $Q_i^{OE}$  in generation  $i$ , and most important  $f(E, \sigma^2)$  is 'some function of the variance in offspring numbers for a given type,  $\sigma^2$ , and of the pattern of variation' (Brandon [1990], p. 20). 'Some function' here must be understood as 'some function or other; we know not what in advance of examining the case'. Moreover, we will have to add to variance other factors that determine the function, such as Beatty and Finsen's skew, or the conditions which Ekbohm, Fagerstrom and Agren identify as making higher variance adaptive, etc. Thus, to be correct, even as a schematic expression, the final term in Brandon's definition will have to be expanded to  $f(E, \sigma^2, \dots)$ , where the ellipsis indicates the additional statistical factors which sometimes combine with or cancel the variance to determine fitness-levels.

But how many such factors are there, and when do they play a non-zero role in fitness? The answer is that the number of such factors is probably indefinitely large, and the reason is given by a fact about natural selection recognized by Darwin and his successors. This fact about selection which

fates our ‘definition’ to being either forever schematic or incomplete is the ‘arms-race’ strategic character of evolutionary interaction. Since every strategy for enhancing reproductive fitness (including how many offspring to have in a given environment) calls forth a counter-strategy among competing organisms (which may undercut the initial reproductive strategy), the number of conditions covered by our *ceteris paribus* clause, or equivalently the number of places in the function  $f(E, \sigma^2, \dots)$  is equal to the number of strategies and counter-strategies of reproduction available in an environment.<sup>9</sup>

In each particular selective scenario, a different specification of Brandon’s definition,  $A^*(E, O)$ , figures in the antecedent of different versions of the  $PNS_{pop}$  and  $PNS_{ind}$ . Properly restricted to the right function  $f(E, \sigma^2, \dots)$  and the right set of statistical features of its reproductive rate for a given environment, each of these versions of the PNSs will presumably be a nomological generalization about natural selection for a given population in a given environment. And the set of these narrowly specific PNSs (each different in the subject matter and the functional form of its antecedent’s fitness-measure) will disjunctively constitute a general principle of natural selections for populations and/or individuals.

The notion that there is no single  $PNS_{pop}$  or  $PNS_{ind}$ , but a family of them, each with a restricted range of application, will be attractive to those biologists uncomfortable with a single principle or law of natural selection, and to those philosophers of science who treat the theory of natural selection as a class of models (Beatty [1981], Lloyd [1994], Thompson [1989]). But these generalizations have mathematically similar antecedents and identical consequents. One will want to ask what they have in common? It cannot be that they share an abstract schema, since abstract schemata do not explain. Could it be that each one is an equally fundamental principle of the theory of natural selection? We believe that these restricted generalizations do in fact have something in common that unifies and explains them, for each of the members of the set of functions  $f_1(E, \sigma^2, \dots)$ ,  $f_2(E, \sigma^2, \dots)$ ,  $f_3(E, \sigma^2, \dots)$  ... measures the same thing—comparative fitness—and identifies it as the cause of the probabilistic claim in each of the generalizations consequents.

And what is comparative fitness, as opposed to its effects in reproduction which measure it? One possible answer is:

a is fitter than b in E = a’s traits result in its solving the design problems set by E more fully than b’s traits.

<sup>9</sup> That this number of strategies and counter-strategies may be indefinitely large forms a crucial component of Rosenberg’s ([2001]) argument that there are no biological laws beyond the PNSs, and their deductive consequences.

This formula (or any of its terminological equivalents) provides a definition of what we label 'ecological fitness', which supervenes on all those relations between an individual and its environment that contribute to the individual's success.<sup>10</sup> Fitness as design-problem solution is, however, famously unattractive to philosophers and biologists (see for instance Lewontin [1978], pp. 257–65). The problems vexing this definition include at least the following ones: (a) it is not obvious how to individuate and count distinct design problems; (b) nor is it clear how to measure the degree to which they are solved by individual organisms; (c) aggregating solutions into an overall level of fitness is difficult in the absence of a common unit to measure ecological fitness; (d) comparing con-specifics which solve different problems to differing extents is equally perplexing. 'x solves more design problems than y' is at least as recalcitrant to measurement as 'x is fitter than y'. Besides the difficulties facing any attempt to operationalize the concept of ecological fitness, there is the objection to its suggestion of teleology in the notion of a 'design problem', and the definition's consequent vulnerability to charges of Panglossian adaptationalism.

It is apparently cold philosophical comfort to defend the design-problem solution definition of fitness by arguing that this litany of difficulties trades on the assimilation of the meaning of a term to its measurement, and fails to recognize the theoretical character of the concept of fitness. Objections to this definition are unlikely to be answered by pointing out that definitions have to stop somewhere, that the definition of a theoretical term must be distinct from the operational measure of the property it names, and that testability is not a matter of theory meeting data one proposition, still less one term, at a time.<sup>11</sup> Or at least none of these considerations have convinced philosophers of biology to give up the project of defining fitness in terms of its effects.

Perhaps the most serious obstacle to accepting the ecological fitness concept is that it is impossible to reconcile with the 'central tendencies' account of the claims of the theory of natural selection now so widely endorsed, for ecological fitness is a relationship between organisms taken two at a time, not a statistical property of populations. Thus, there is among exponents of the 'central tendencies' approach a strong incentive to deal with the problem of defining 'fitness' by simply expunging the concept altogether from the theory of natural selection. No fitness, no fitness problems. This strategy is adopted explicitly by Matthen and Ariew ([2002]).

<sup>10</sup> We adapt this notion from Brandon ([1990]) and Endler ([1986]), who describe a similar notion of within-generation success, distinct from a concept of fitness as transgenerational (i.e. reproductive) success. As noted below, Matthen and Ariew ([2002]) label this notion 'vernacular fitness' and seek to expunge it from the theory of evolution altogether.

<sup>11</sup> Cf. Rosenberg ([1983]).

But, as we will now show, expunging ecological fitness from the theory of natural selection makes the theory unrecognizable. This means that despite its measurement-problems, the ecological fitness concept, whether or not it must ultimately be understood in terms of the solution to design problems, turns out to be indispensable to the theory of natural selection.

### **The PNS and the second law of thermodynamics**

At least since the work of Peirce,<sup>12</sup> philosophers have been trying to understand the claims of the theory of natural selection by treating it on analogy with the second law of thermodynamics. Matthen and Ariew ([2002]) write, for instance: ‘As Fisher kept emphasizing, it is statistical thermodynamics—not Newtonian dynamics—that provides the closest parallel in physics to the theory of natural selection’ (p. 72). Philosophers seeking to treat the theory of natural selection as a claim about central tendencies exclusively have reason to pursue this similarity, for (a) both the  $PNS_{pop}$  and the second law of thermodynamics have probabilistic consequents not open to interpretation as subjective degrees of belief or probabilistic propensities, and (b) the second law is a regularity about ensembles, not the individuals out of which they are composed.

We may state the second law of thermodynamics as follows:

(x)(y) [x, y are states of a closed thermodynamic system and y is later than x  
 → Probably (the entropy of y is greater than the entropy of x)]

The two PNSs have a probabilistic consequent isomorphic to the second law’s consequent:

→ Probably (x’s size is larger than y in E at some generation n’ later)

It is this similarity in probabilistic consequents that seems to have encouraged philosophers to treat the PNS as a claim about ensembles, like the second law, and to treat fitness as a property of ensembles, on a par with ‘entropy’.<sup>13</sup>

<sup>12</sup> In ‘Fixation of Belief’ ([1877], Section 1), Peirce writes: ‘Mr. Darwin proposed to apply the statistical method to biology. The same thing had been done in a widely different branch of science, the theory of gases. Though unable to say what the movements of any particular molecule of gas would be on a certain hypothesis regarding the constitution of this class of bodies, Clausius and Maxwell were yet able, by the application of the doctrine of probabilities, to predict that in the long run such and such a proportion of the molecules would, under given circumstances, acquire such and such velocities; that there would take place, every second, such and such a number of collisions, etc.; and from these propositions were able to deduce certain properties of gases, especially in regard to their heat-relations. In like manner, Darwin, while unable to say what the operation of variation and natural selection in any individual case will be, demonstrates that in the long run they will adapt animals to their circumstances.’

<sup>13</sup> Besides Matthen and Ariew ([2002]), another recent example of this trend is the related work of Walsh, Lewens and Ariew ([2002]). In the latter, see p. 463, and p. 463 note 6, for example.

The trouble with the analogy between the  $PNS_{pop}$  and the second law of thermodynamics is that the features that make for the emergent mysteries of the second law are largely absent from the foundations of the theory of natural selection. Once we understand the differences between entropy and fitness, the temptation to treat the theory of natural selection as a claim solely about ensembles disappears.

The emergent character of the second law is generated by the fact that entropy is not a property of the individual components of an ensemble, but of the ensemble as a whole. The standard explanation of how entropy emerges from the behavior of the members of the ensemble remains highly problematical.

To see why, consider the simplest case in which a thermodynamic system—say a quantity of a gas in a container—is treated as an ensemble of particles moving in accordance with Newtonian dynamical laws. Following Albert ([2000], p. 43ff.), call a specification of which particles are where in the container, and what their specific momenta are, an ‘arrangement’, and a specification of how many particles are within a given region of the container and a given range of momenta a ‘distribution’. The entropy of the system depends on the distribution of the particles and not the particular arrangement of them. Any one distribution is of course compatible with more than one arrangement of particles. The particles change position and momenta in accordance with deterministic Newtonian laws, and the number of physically possible arrangements of particles that realize any one distribution increases as the particles spread out in space and in momentum values. The increase in entropy that the second law reports is a result of this fact about arrangements and distributions: in the long run, later distributions supervene on a larger number of arrangements than earlier ones do. The larger the number of arrangements for a given distribution, the higher the entropy. Entropy is thus accounted for in terms of Newtonian concepts of position and momentum via the concepts of distribution and arrangement.

The flaw in this story is that we have no right to hold that the number of arrangements at the earlier time is less than the number of arrangements at the later time. Since Newtonian momentum and space-time location can take on a continuum of values, the number of arrangements compatible with (almost) any single distribution is infinite, and there is no unique way to measure the size of these infinities. Within any given region of space and range of momentum values for any one particle, the position and momentum of the particle can take up a continuum of values. If the earlier ‘smaller’ number of arrangements compatible with a given distribution is infinite in number, and the later, larger ‘number’ of arrangements is also infinite in number, we cannot appeal to differences in the number of arrangements on which given distributions supervene to explain the increase in entropy that the second law reports.

Thus, both entropy as a property and the second law as a regularity are said to be irreducible ensemble-level matters.<sup>14</sup>

But the theory of natural selection is not vexed by the problems that bedevil a reduction of thermodynamic properties to Newtonian dynamics, which make entropy an emergent property of an ensemble, and prevent us from turning the schematic derivation of the second law into a complete explanation. In evolutionary theory, all we need in order to understand where the fitness-coefficients of populations come from is the ‘concession’ that there is such a thing as comparative differences in (ecological) fitness between pairs of individual organisms; and that these differences can be aggregated into fitness differences between populations.

Recall the distinction drawn between the  $PNS_{pop}$  and the  $PNS_{ind}$  in Section 1 above. Treat fitness as it figures in both PNSs as a matter of solving design problems (measured by some demographic statistic). Then the truth of the  $PNS_{pop}$  follows from the truth of the  $PNS_{ind}$  by simple arithmetical aggregation. There is no difficulty explaining where ‘comparative fitness’ in the  $PNS_{pop}$  ‘comes from’: it is just the average over the compared populations of the comparative fitnesses of the individual members of the populations. There is nothing at the ensemble level here emergent from the properties at the individual level the way there is in thermodynamics. There is no new property of the whole ensemble—such as entropy—utterly dissimilar from any properties at the level of the components of the ensemble. There is just the average of actual comparative fitness relations among pairs of organisms. It is true that *measuring* comparative fitness as it figures in the  $PNS_{pop}$  and the  $PNS_{ind}$  is a matter that moves in the opposite direction from the direction of explanation as it obtains between these principles. That is, to get a quantitative handle on the degree to which one organism solves the design problems set by the environment more fully than another, one must aggregate over like creatures, whence the attractions of the probabilistic propensity ‘definition’—or rather one or another of its disjuncts—to measure values of ecological fitness. When this requires actually collecting data about reproduction rates, variances in them, skews, etc., over multiple generations, independent evidence for the explanatory role of the  $PNS_{ind}$  is rendered invisible.

As noted, the PNSs all do share with the second law of thermodynamics a probabilistic operator in their consequents. But this probabilistic operator is not the feature of the second law that obscures its foundations in Newtonian dynamics. The distinctive problem of the second law is that we would like to be able to say that states of higher entropy of an ensemble depend on

<sup>14</sup> The apparent insolubility of this problem of reductively explaining thermodynamics has been diagnosed by Lawrence Sklar in several works ([1993], [1999]). Sklar concludes that we must resign ourselves to building these probabilities into thermodynamics at its own level.

distributions which are realized by large number of arrangements of its components. We cannot say this, because every distribution includes an infinite number of physically possible arrangements and there is no non-arbitrary measure on these infinities that will enable us to compare their size. This problem for thermodynamics, of identifying a measure on infinite sets of different cardinalities, simply does not occur in the theory of natural selection. The *fitness* of an ensemble is just nothing like the *entropy* of an ensemble, just because *unlike* entropy, fitness is a calculable value of the properties of the components of the ensemble.

### 3 Drift and selection

There is in fact a significant parallel between the  $\text{PNS}_{\text{pop}}$  and the second law, but it does not substantiate the conclusion that the former is, like the latter, a law about irreducible ensembles. The significant parallel between the PNSs and the second law is to be found in the probability operators in their consequents. It is this probability concept which makes ecological fitness indispensable to the theory of natural selection's claims about ensembles and populations, as we now show.

The probabilistic character of the consequent of the PNSs is what makes room for drift. If the long-run relative frequency of *a*'s having more offspring than *b* is greater than the long-run relative frequency of *b*'s having more offspring than *a*, then this frequency is compatible with any actual finite frequency. When finite actual frequencies approach the long-run relative frequency cited in the PNSs, the principles explain these finite actual frequencies. When the finite actual frequencies do not approach the long-run frequencies, the alternative explanations are (a) the PNS is false, or (b) the divergence between the long-run and the actual frequencies is a matter of drift. Exclude the first alternative. As we shall see, drift plays its role in natural selection only against the background of disaggregated pairwise ecological fitness differences among individual biological entities that cause differential reproduction.

Despite the heavy weather made of it in the philosophy of biology, drift is perfectly easy to understand. Consider everyone's favorite example: coin tossing. A fair coin has a long-run relative frequency of coming up heads equal to .5. When tossed 1000 times in batches of ten, it comes up heads, say, a total of 491 times, but in some of the batches, it will come up heads 6, 7 or even 8 times. The (weak) law of large numbers tells us that if the long-run relative frequency of heads is .5 then the subjective probability of the actual frequencies approaching .5 converges on 1.0 as the number of coin flips increases. By contraposition, as the actual number of fair coin flips decreases, the subjective probability that the actual frequency of heads equals .5 will decrease. It is fallacious to infer from the law of large numbers, a theorem of

the calculus of probabilities, that actual frequencies approach the long-run relative frequency as the number of tosses grows larger. It is equally fallacious to infer that the failure of actual frequencies to approach the long-run frequencies shows that the coin is not fair.

The causal explanation of the divergence of a finite sample from the long-run relative frequency of coin tosses is to be sought in the fact that the initial conditions of the actual coin tosses were not representative of the set of initial conditions that give rise to the long run. To see this, imagine a spring-loaded apparatus for tossing quarter-sized disks, and a single physically bilaterally symmetrical quarter-sized disk marked H and T, such that whenever the disk sits in the apparatus with H up, and the spring is released, the disk is shot out on a single parabolic trajectory with three rotations of the disk that always results in its landing H-side up (and vice versa if its starts out T-face up in the apparatus).

There is nothing counterfactual about this physical system. It is a deterministic one<sup>15</sup> in which all the actual sequences of H-flips come up 100 % H, and similarly for T-flips; obviously, so long as the spring retains its elasticity, the disks are not worn, etc., the long-run relative frequency  $P$  (the disk comes up H on landing/the disk is H-side up in apparatus) = 1. Now, consider a real quarter, and a real thumb-and-forefinger coin-flipping 'device'. This physical system does not differ from our machine-and-disk system in any physically relevant way. Accordingly, it must also be a deterministic system. But, when the quarter is flipped head-side up, say 100 times, it lands heads 47 times and tails 53 times, and when it is flipped 1000 times, it comes up heads 502 times, and so on. We infer that the long-run relative frequency  $P$  (the quarter comes up H on landing/the quarter is H-side up on the forefinger) = .5, and we know perfectly well where this probability 'comes from'. It is the result of the fact that the initial conditions of the coin flipping which deterministically bring about an outcome of H or T in each case are distributed into two sets. One of these sets of initial conditions together with the relevant Newtonian laws determines a set of paths from thumb to table top which results in *heads*, while the other set of initial conditions together with the same set of laws determines paths to the table top resulting in *tails*. If there were 47 heads out of 100 tosses then there were 47 initial conditions in the former set. If it is a fact that as the number of tosses increases, the number of initial conditions in the heads-outcome set approaches 50%, then the number of heads-outcomes

<sup>15</sup> That is, it is deterministic with the qualification given in Rosenberg ([2001], p. 537): 'the actual world, which is quantum-indeterministic in its fundamental laws of working, asymptotically approaches Newtonian determinism for objects as large as our coin-tossing device owing to the fact that the probabilities of violation of Newton's laws by macroscopic objects are so low that there is not a single actual violation in the amount of time taken up by the whole history of the [actual] world.'

approaches 50%. When ratio of heads to tails varies from exactly 50:50 we can be sure that the cause is that the distribution of initial conditions is not 50:50. Thus, when an actual series results in 50% heads, the explanation is that 50% of the initial conditions were of the heads-resulting sort, and when the actual series is not 50:50, the explanation is that the initial conditions were not distributed 50:50.

Compare the case of a set of 100 uranium atoms each with a 50% chance of emitting an alpha particle in a period of time  $t$ . If only 47 atoms emit alpha particles, there is no reason to assign a cause in the initial conditions realized by those 47 uranium atoms, for alpha particle decay is a fundamentally indeterministic process. The initial conditions of those 47 atoms do not differ from the initial conditions of the 53 atoms which did not emit alpha particles in the time period in question. And there is no explanation of why 47 of the atoms emitted alpha particles and 53 of them did not.

Suppose we have evidence that the set of initial conditions of a real series of coin flips is divisible into two equal sets—one of which results in H and the other T. This evidence will consist in the bilateral symmetry of the coin, the inability of the flipper to control initial conditions very accurately, etc. And suppose that the series of flips results in 50 H and 50 T. Well, then, the explanation is the equal size of the two sets of realized initial conditions. Suppose that among the set, however, the 20<sup>th</sup> through 23<sup>rd</sup> tosses were 4 consecutive Heads. This is an improbable event,  $P(H,H,H,H) = .0625$ , and not explainable by appeal to the equal distribution of initial conditions into H-resulting and T-resulting sets. It is explained by showing how the initial conditions in tosses 20 through 23 together with Newton's laws resulted in Hs. It is certainly true that in the long run when the initial conditions are equally distributed between heads-resulting and tails-resulting initial conditions, 6.25 % of the time 4 consecutive heads comes up. But this is either no explanation of why 4 heads came up when they did, on tosses 20–23, or only a small part of the explanation, or an explanation of something else (viz, that 6.25 % of large numbers of fair-coin tosses result in 4 consecutive heads), or an explanation that satisfies very unstringent standards on explanatory adequacy. By contrast, in alpha particle emission among uranium atoms, that 4 contiguous atoms emitted alpha particles in the same period when each had only a 50% probability of doing so is maximally explained by the calculation that there was an objective and not further explainable 6.25 % chance of its happening to every 4 contiguous uranium atoms.

When we are presented with various actual sequences of Hs and Ts, we frame explanations of them which vary in the stringency of the conditions on explanatory adequacy which they are expected to meet as a function of our interest in particular series of outcomes. Usually, our interests in the details are so weak that we are satisfied with an explanation for why a particular

series of Hs and Ts approaches a 50:50 ratio that appeals to a division of initial conditions into sets of sizes that approach 50:50. The role in the explanans of the premise that the initial conditions of the coin-tosses are equally divided between those that result in heads and those that result in tails is just a special case of the appeal to randomness in an experimental treatment. The empirical generalization which explains why the coin-tossing ratio approaches 50:50 tells us about what happens if a random trial is repeated over and over, independently and under conditions otherwise identical: under these conditions, the fraction of trials that result in a given outcome converges to a limit as the number of trials grows without bound.

Two things to note. First, when, as in the case of four consecutive heads, the fraction of trials does not converge, it follows from the empirical generalization mentioned above that the trials are not random, or not independent, or conditions have changed. And these facts must take part in the explanation of the four consecutive heads.<sup>16</sup> More important, the explanation of why large numbers of tosses of fair coins approach 50:50 relies on the *randomness* of the trials.

What does randomness consist in when it comes to coin-flipping? Randomness consists in each of the physically possible initial conditions of a coin-flipping system being *equiprobable* (whence the equality of the number of initial conditions resulting in heads and in tails). Since coin-flipping is an asymptotically deterministic affair, the source of the equiprobable randomness cannot be anything like the probabilistic propensities resulting from quantum processes. And while it may be reasonable, *ceteris paribus*, to adopt subjective probabilities or betting odds that are the same for all possible initial conditions of coin-flipping, the equal distribution of all physically possible heads-causing and tails-causing initial conditions does not turn on anyone's epistemic states. It seems to be a fact about the world independent of subjective probabilities and betting odds that in the long run the physically possible initial conditions of fair-coin tossing are equiprobable.<sup>17</sup>

<sup>16</sup> For this reason, we reject Matthen and Ariew's ([2002], p. 61) claim to the contrary that the same statistical facts necessarily explain the case of two heads out of four tosses *and* four heads out of four tosses of a fair coin.

<sup>17</sup> Here (unlike the entropy/fitness disanalogy) we do have the same problem that vexes the long-run relative frequency probability operator in the consequent of the second law—the claim that probably entropy will increase. As with fair coin flipping, we need to assume that all the actual dynamic states of the constituents of the ensemble are distributed equally into all the possible dynamic states. But no one in physics or its philosophy, from Gibbs to Sklar, has been able to ground the assumption of equiprobability to general satisfaction. The situation is no different in coin flipping. The claim that all the possible initial conditions are equiprobable might well be called a metaphysical commitment.

What is the bearing of all this on the PNSs? Fitness differences are much more like coin biases than they are like differences in alpha particle emission.<sup>18</sup> Suppose we have a heads-biased coin—one biased because asymmetrical in shape, density, magnetic charge, etc. This coin comes up heads with a long-run relative frequency of .7, when flipped often enough by a given thumb-and-forefinger apparatus on random independent trials, i.e. when the initial conditions of flipping are equiprobable. Evolutionary fitness differences have the same consequences as coin biases. If an organism of type **a** has a fitness coefficient of 1 and an organism of type **b** has a fitness coefficient of .4285, then as a matter of long-run relative frequency, the **a**-type will have 7 offspring to **b**-type's 3 offspring, just as a coin biased .7 to heads will, in the long run, come up 7 heads for every 3 tails. Assuming that selection is an asymptotically deterministic process that differs only by degree from coin-flipping, fitness differences will have results of the same character as tossing biased coins has. It will be an empirical fact that when initial conditions are random and trials are independent, actual frequencies of .7-biased coin-flips approach the long-run relative frequency of 7:3 as they increase in size, and actual numbers of offspring of organisms whose fitness ratios are 7:3 will approach the long-run relative frequency of 7:3 in offspring numbers.

In both cases, divergence from the 7:3 ratio will be deemed to be drift, in retrospect at least, if the divergence declines as numbers of tosses or generations increases. And each divergence will be in principle explainable deterministically by identifying their initial conditions. The explanation of the divergence will presumably show that the divergence does not disconfirm the long-run relative frequency hypothesis, as the initial conditions in the divergence were rare, improbable and unrepresentative of the whole population of initial conditions. In practice of course, these initial conditions are not in fact epistemically accessible either before or after the events in the divergent series (this is what makes coin-flipping a useful device for gambling of course).

How do we decide whether a divergence from a long-run relative frequency prediction about fitness differences is a matter of drift, a disconfirmation of the hypothesis of natural selection, or a reflection of a mismeasurement of fitness differences to begin with? Suppose we measure the fitness differences between population **a** and population **b** to be in the ratio of 7:3, and suppose further that in some generation, the actual offspring ratio is 5:5. There are four alternatives: (a) the fitness measure of 7:3 is right but there was drift—i.e. the initial condition at this generation are unrepresentative of those which obtain

<sup>18</sup> Of course some fitness differences may supervene on quantum indeterministic processes, as acknowledged in Section 1 above. See also Stamos ([2001]), Glymour ([2001]) and Rosenberg ([2001]), and footnote 10 above. The probabilistic outcomes of fitness differences once established cannot be due solely to quantum indeterministic processes, or they would be much less frequent than in fact they are.

in all relevant generations; (b) the fitness measure of 7:3 was incorrect and there was no drift; (c) there was drift, and fitness was wrongly measured (d) the principle of natural selection is disconfirmed. How do we discriminate among the first three of these four alternatives? The answer is critical to seeing the role of ecological fitness in the theory of natural selection.

In the absence of information about the initial conditions of the divergence, there is only one way empirically to choose between the three alternatives (a)–(c). This way requires access to ecological fitness differences. This access we have, at least in principle, when we make comparisons between the degree to which compared individuals solve specified design problems which biologists identify. These comparisons give the independent empirical content to the notion of ecological fitness, while allowing for it to be (fallibly) measured by probabilistic propensities to leave offspring. For example, we can tell that white-coated arctic prey are fitter than their dark-coated competitors since they have solved a pressing design problem better. We can make this fitness-judgment without counting offspring, though barring drift we expect such head-counts to measure the ecological fitness difference instantiated. If the theory adverts to ecological fitness differences, it has the resources, at least in principle, to decide whether the divergence from predicted long-run relative frequencies, especially where small populations are concerned, is a matter of drift or selection, i.e. whether demographic changes are due to ecological fitness differences or the unrepresentativeness of the initial conditions of individual births, deaths and reproductions.

The problem of distinguishing drift from selection in ensembles (large populations) has the same character, and is in principle susceptible to the same solution. We can distinguish drift from selection in ensembles *if* we accept that there is such a thing as ecological fitness differences, *if* we have at least in principle access to the initial conditions of births, deaths and reproductions, taken one at a time, and *if* we accept that these individual differences aggregate into ensemble differences. The fact that the solution is often available only in principle, and not to be obtained in practice, is reflected in our willingness to be satisfied by explanations that pass only the lowest of stringency tests. But at least in principle, in these cases there must be a causal explanation of the individual fitness differences; for without it we cannot distinguish drift from selection among ensembles, and the combination of both (which always obtains since populations are not infinite and no actual run is a long run) from the falsity of the theory of natural selection altogether. Because there is always some drift, there is in the end no substitute for ecological fitness and no way to dispense with its services to the theory of natural selection. And since ecological fitness is ultimately a relationship between organisms taken two at a time, the theory is as much a set of claims about individuals as it is about ensembles. Moreover, since fitness is

ecological, it must be distinguished from ‘expected reproduction rates’. This result thankfully frees us to treat selection as a contingent causal process in which individual fitness differences are the causes and subsequent population differences are the effects.

#### **4 Conclusion**

Biologists and philosophers who seek an understanding of the theory of natural selection and its application to the natural history of this planet require a concept of ecological fitness. If the best way to define this term is by way of the notion of overall design-problem solution, then biologists and philosophers will have to decide if they can live with such a definition, despite its teleological suggestion and its measurement difficulties. Like democracy, fitness as design-problem solution may turn out to be the best among a large collection of unsatisfactory alternatives. If this is right, biologists and philosophers need to re-examine how fitness can be ‘measured’ in ways independent of offspring contribution. Leaving this matter open, we need at least to see that attempts to define fitness solely as a probabilistic propensity are unavailing for biological reasons, and attempts to treat the concept as a property of ensembles, along the lines of ‘entropy’, obscure fundamental differences between fitness and entropy. And finally, the evolutionary contrast between selection and random drift makes indispensable a causal concept of ecological fitness.

#### **Acknowledgements**

Thanks to Robert Brandon, Elliott Sober, Dan McShea and Stefan Linquist and an anonymous reviewer, for helpful comments and suggestions. We also profited from discussions with the members of the Duke Center for Philosophy of Biology. Special thanks also to Mohan Matthen, André Ariew and Dennis Walsh for clarifying remarks. Bouchard received support from the Social Sciences and Humanities Research Council of Canada.

*Center for Philosophy of Biology  
Box 90743  
Duke University  
Durham  
NC 27708  
USA  
alexrose@duke.edu*

## References

- Albert, David, [2000]: *Time and Chance*, Cambridge, MA: Harvard University Press.
- Ariew, A. and Lewontin, R. 2004: 'The Confusions of Fitness', *British Journal for the Philosophy of Science*, **55**, pp. 347–363.
- Beatty, J. [1981]: 'What's Wrong with the Received View of Evolutionary Theory?', *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* vol. 1980, v.2, pp. 397–426.
- Beatty, J. and Finsen, S. [1987]: 'Rethinking the Propensity Interpretation', in M. Ruse (ed.), *What Philosophy of Biology Is*, Boston: Kluwer.
- Beatty, J. and Mills, S. [1979]: 'The Propensity Definition of Fitness', *Philosophy of Science*, **46**, pp. 263–88.
- Bossert, W. and Wilson, E. O. [1974]: *A Primer of Population Biology*, Stanford: Sinaur.
- Brandon, R. [1990]: *Adaptation and Environment*, Princeton, NJ: Princeton University Press.
- Brandon, R. [1978]: 'Adaptation and Evolutionary Theory', *Studies in the History and Philosophy of Science*, **9**, pp. 181–206.
- Brandon, R. and Carson, S. [1996]: 'The Indeterministic Character of Evolutionary Theory', *Philosophy of Science*, **63**, pp. 315–37.
- Dawkins, R. [1986]: *The Blind Watchmaker*, New York: W. W. Norton.
- Earman, John [1986]: 'A primer on Determinism', University of western ontario series in Philosophy of Science, v. 32. Dordrecht, Boston, Norwell, MA, USA: D. Reidel Pub. Co. Kluwer Academic.
- Ekbohm, G., Fagerstrom, T., and Agren, G. [1980]: 'Natural Selection for Variation in Offspring Numbers: Comments on a Paper by J. H. Gillespie', *American Naturalist*, **115**, pp. 445–7.
- Endler, J. [1986]: *Natural Selection in the Wild*. Princeton, NJ: Princeton University Press.
- Fisher, R. A. [1930]: *The Genetical Theory of Natural Selection*. Oxford: The Clarendon Press.
- Gillespie, G. H. [1977]: 'Natural Selection for Variances in Offspring Numbers: A New Evolutionary Principle', *American Naturalist*, **111**, pp. 1010–4.
- Glymour, B. [2001]: 'Selection, Indeterminism and Evolutionary Theory', *Philosophy of Science*, **68**, pp. 518–35.
- Lewis, D. [1984]: *Philosophical Papers*, Vol. 1, Cambridge, MA: Harvard University Press.
- Lewontin, R. [1978]: 'Adaptation', *Scientific American*, 239, no.3, pp. 156–69.
- Lewontin, R. C. [1970]: "The Units of Selection", *Annual Review of Ecology and systematics*, **1**, 1–14.
- Lewontin, R. C. [1974]: *The Genetic Basis of Evolutionary Change*, New York: columbia University Press.
- Li C. C. [1995]: *Population Genetics*, Chicago: University of Chicago Press.
- Lloyd, E. [1994]: *The Structure and Confirmation of Evolutionary Theory*, Princeton, NJ: Princeton University Press.

- Matthen, M. and Ariew, A. [2002]: 'Two Ways of Thinking about Fitness and Natural Selection', *Journal of Philosophy*, **99**, pp. 55–83.
- Peirce, C. S. [1877]: 'The Fixation of Belief', *Popular Science Monthly*, **12**, pp. 1–15.
- Railton, P. [1978]: 'A Deductive Nomological Model of Probabilistic Explanation', *Philosophy of Science*, **45**, pp. 206–26.
- Rosenberg, A. [1983]: 'Fitness', *Journal of Philosophy*, **80**, pp. 457–73.
- Rosenberg, A. [2001]: 'Indeterminism, Probability and Randomness in Evolutionary Theory', *Philosophy of Science*, **68**, pp. 536–44.
- Rosenberg [2001]: "How is Biological Explanation Possible", *British Journal for the Philosophy of Science*, **52**, pp. 735–760.
- Salmon, W. [1966]: *Foundations of Scientific Inference*, Pittsburgh: University of Pittsburgh Press.
- Sklar, L. [1993]: *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*, New York: Cambridge University Press.
- Sklar, L. [1999]: 'The Reduction(?) of Thermodynamics to Statistical Mechanics', *Philosophical Studies*, **95**, pp. 187–202.
- Sober, E. [1984]: *The Nature of Selection*, Cambridge, MA: MIT Press.
- Sober, E. [1993, 2000]: *The Philosophy of Biology*, Boulder, CO: Westview & Harper Collins.
- Sober, E. and Wilson, D. S. [1998]: *Unto Others: The Evolution and Psychology of Unselfish Behavior*, Cambridge, MA: MIT Press.
- Sober, E. [2002]: 'The Two Faces of Fitness', in *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*, Cambridge: Cambridge University Press.
- Stamos, D. [2001]: 'Quantum Indeterminism and Evolutionary Biology', *Philosophy of Science*, **68**, pp. 164–84.
- Sterelny, K. and Kitcher, P. [1988]: 'Return of the Gene', *Journal of Philosophy*, **85**, pp. 339–62.
- Strickberger, M. [1985]: *Genetics*, 3rd edn, New York: McMillan.
- Thoday, J. M. [1953]: 'Components of Fitness', in *Symposium of the Society for Experimental Biology: Evolution*, **7**, pp. 96–113.
- Thompson, P. [1989]: *The Structure of Biological Theories*, New York: SUNY Press.
- Walsh, D., Lewens, T. and Ariew, A. [2002]: 'The Trials of Life: Natural Selection and Random Drift', *Philosophy of Science*, **69**, pp. 452–73.