Brit. J. Phil. Sci. **64** (2013), 851–881

A New Foundation for the Propensity Interpretation of Fitness

Charles H. Pence and Grant Ramsey

ABSTRACT

The propensity interpretation of fitness (PIF) is commonly taken to be subject to a set of simple counterexamples. We argue that three of the most important of these are not counterexamples to the PIF itself, but only to the traditional mathematical model of this propensity: fitness as expected number of offspring. They fail to demonstrate that a new mathematical model of the PIF could not succeed where this older model fails. We then propose a new formalization of the PIF that avoids these (and other) counterexamples. By producing a counterexample-free model of the PIF, we call into question one of the primary motivations for adopting the statisticalist interpretation of fitness. In addition, this new model has the benefit of being more closely allied with contemporary mathematical biology than the traditional model of the PIF.

doi:10.1093/bjps/axs037

Advance Access published on March 5, 2013

1 Introduction

The propensity interpretation of fitness (PIF) was introduced in the late 1970s with two articles, one by Robert Brandon ([1978]) and the other by Susan Finsen (née Mills) and John Beatty ([1979]). These papers, among other things, provided a solution to what has since come to be known as the 'tautology problem': if fitness is defined in terms of actual evolutionary outcomes, fitness cannot then causally explain these outcomes—the 'survival of the fittest' reduces to 'the survival of those that survive'. These papers proposed that fitness (or, in Brandon's terminology, 'adaptedness') is a probabilistic propensity to produce offspring. Each organism, that is, is taken to have a disposition to produce certain numbers of offspring, with differing probabilities associated with each possible offspring number. If the fitness of an individual organism is traceable to this propensity, then it is clear that the tautology problem is solved. The fittest organisms will, indeed, tend to survive more often than their less fit counterparts, just as a sturdier glass will tend to break less often than a fragile one.

Fitness, however, fills more roles than merely the prevention of tautology. Most models of evolutionary change employ fitness as a scalar numerical value, comparable between organisms. In addition to providing a rank ordering of the organisms in a population—which can justify claims like '*a* is fitter than *b*'—these fitness values are utilized by models such as those in population genetics to predict the future evolutionary trajectory of a given population.

The PIF, then, has traditionally been presented alongside a mathematical model which can serve to translate this probability distribution into a single, privileged measure on the distribution. The primary such mathematical model of the PIF, introduced by Brandon, has defined fitness as an organism's expected or average number of offspring, weighted by the associated probabilities. He described this formally as

$$A(O,E) = \sum P(Q_i^{OE})Q_i^{OE}, \tag{1}$$

where each Q_i^{OE} is a possible number of offspring and $P(Q_i^{OE})$ is the probability of that number of offspring being realized. As mentioned above, Brandon used the term 'adaptedness' for fitness and $A(O,E)$ should thus be read as 'the

adaptedness (i.e. fitness) of organism *O* in environment *E*'. This is the standard model of the PIF, and it is shared across most major presentations of the PIF, including those of Mills and Beatty ([1979]), Beatty and Finsen ([1989]), Brandon ([1990]), and Sober ([2001]).

1.1 The 'Generality Problem'

When the PIF was introduced, Brandon also saw that it could be useful for the solution of another problem in the philosophy of biology, one which we will call here the 'generality problem'. One area of work in the philosophy of biology has endeavored to theorize at a very abstract level about the process of evolution by natural selection—we might, consistent with similar terminology in the philosophy of physics, call this the study of the 'foundations' of evolutionary theory. Explanations of evolution at this level do not focus on particular episodes of natural selection, but rather on what it is that is common to every instance of natural selection, across every environment, system of heredity, unit of selection, and so on where natural selection might be instantiated.

Returning to Brandon, in the same paper in which the PIF is introduced, he introduces the following as a 'law of nature' (which he would later call the principle of natural selection (Brandon [1990], p. 11):

> If *a* is better adapted than *b* in environment *E*, then (probably) *a* will have more (sufficiently similar) offspring than *b* in *E*. (Brandon [1978], p. 187)

He argues that it is this law, or something like it, that is presupposed by all general explanations of natural selection, including the three traditional Lewontin conditions for evolution by natural selection: variation, heritability, and fitness differences (Lewontin [1970]).

It is thus crucial to the understanding of philosophical work such as this that we provide a definition of what Brandon calls 'better adapted' and which many other authors, including Bouchard and Rosenberg ([2004]), call 'fitter'. That is, we need a notion of fitness that is capable of serving in the phrase '*a* is fitter than *b*' in *every* evolutionary system in which a statement like Brandon's principle of natural selection (PNS) holds true. This is what we will call the generality problem: the problem of producing a notion of fitness which may be correctly applied in explanations that encompass all possible evolutionary systems.

The generality problem does not only appear in the context of these unrestricted, general principles of natural selection, however. Indeed, were this its only use, it would be a matter of debate whether such a notion of fitness were really necessary at all, as some authors have constructed frameworks for

natural selection on which a PNS like that deployed by Brandon is not necessary.[1] The reason why the generality problem is so important is that it is also central to the debate over the causal structure of natural selection, fitness, and genetic drift—the debate between the 'statisticalists' and 'causalists'.

When, for example, Abrams ([2007], p. 670) proposes as an 'elaboration of the PIF' that 'if there are individual-level probabilities which are in some sense causal, natural selection and drift are causal in the same sense', and Walsh ([2010], p. 168) argues in response that 'fitness distribution explains but does not cause population change', these authors are not arguing over the causal forces present in some particular biological population. Rather, they are asking us to consider what the appropriate interpretation of evolutionary theory is, again, in every circumstance in which it applies. Does fitness reflect a causal property in all cases? Or is it merely a method of bookkeeping, a subjective tally of objective organismic lives and deaths? Again, it seems that a prerequisite for this debate is to find a concept of fitness and a mathematical model of that concept which apply in all cases—to find, that is, a solution to the generality problem.

And the generality problem is not exclusive to the philosophical domain. Lewontin proposes his 'three principles' for describing evolution, the second of these is that '[d]ifferent phenotypes have different rates of survival and reproduction in different environments (differential fitness)' (Lewontin [1970], p. 1). This invocation of 'differential fitness' is not relativized to any specific biological population, or even any specific model of fitness. Later work in mathematical biology has attempted to rigorize this notion of general fitness. Metz *et al.* ([1992], p. 198) argue that

> [...] the biomathematical literature of the last 10–20 years reflects the implicit acceptance of a common evolutionary framework, the core idea being that there exists a unique general fitness measure that concisely summarizes the overall time course of potential invasions by initially rare mutant phenotypes.

The model of fitness that we will propose here, as we will see later, extends this work of Metz *et al.* to the level of the fitness of individuals and draws out its philosophical implications, unifying the philosophical and biological approaches to the generality problem.

One obvious objection to this project is that it is not clear that we require a general concept of fitness to solve the generality problem. Several authors have endeavored to construct theories of natural selection that make no reference to fitness. Prominent among these is the work of Millstein, who notes that 'selection requires differences in abilities to survive and reproduce'. She goes on

[1] The most prominent recent approach which discards the PNS is the spatial framework of Godfrey-Smith ([2009]).

to say that some would term these ' "fitness differences"—I avoid the term to prevent entanglement in disputes over the definition of fitness' ([2006], p. 643). Elsewhere, she defines selection in general as a 'discriminate sampling process whereby physical differences between organisms are causally relevant to differences in reproductive success' (Millstein [2006], p. 640), substituting causally relevant physical differences in place of fitness differences. If such a definition of selection will suffice, why do we need to produce a general model of fitness at all?

We claim that this picture of selection is perfectly reasonable, but only insofar as it smuggles in an implicit reference to a concept of fitness. It's clear that not just any physical difference, or even any physical difference that's causally connected to survival and reproduction, will suffice for being counted as taking part in selection. Each individual mammal, for example, has a unique pattern of hair follicles, and if hair is causally relevant to survival in some species, then *a fortiori* the pattern of individual hairs is as well. But it does not therefore follow that there is a selective difference between each pair of individuals that is due to their follicle pattern difference. We thus need some way to cash out selection in terms of relevant physical differences between organisms. The causal connection of a physical difference with survival and reproduction works as a basic criterion of relevance, but (as we have seen) seems to occasionally include irrelevant features. Furthermore, the numerical quantification of selection—surely an important task for biology—requires a notion that can differentiate just how relevant a given physical difference is to selection. The elaboration of this measure (quantifying how relevant a given physical difference is to an individual's reproductive success) will, we argue, just consist in the elaboration of a model of individual fitness.

Finally, one more approach to the generality problem should be mentioned here. The 'statisticalist interpretation', as the position of Walsh, Ariew, Matthen, Lewens, and others has come to be known, attempts to solve this problem by replacing the PIF with an interpretation of fitness as a set of facts about the statistical distribution of evolutionary outcomes. Walsh, for example, states that 'what it is for a change in relative trait frequencies to constitute selection (or drift) is merely for it to be susceptible to a certain kind of statistical description' ([2007], p. 282).[2] As we will now see, one of the primary reasons for adopting such a position is that there does not exist an account of the PIF that is free of counterexamples. By producing such an account here, we therefore substantially weaken a key motivation

[2] See Matthen and Ariew ([2002]), Walsh *et al.* ([2002]), Krimbas ([2004]), Brunnander ([2006]), Walsh ([2007]), Ariew and Ernst ([2009]) and Walsh ([2010]) for other works in the statisticalist program.

for adopting the statisticalist position along with their solution to the generality problem.

1.2 Counterexamples to the PIF

In addition to offering their own solution, the statisticalists (as well as some reluctant yet honest propensity theorists) have offered several counterexamples that purport to demonstrate that the PIF is not in fact suitable as an answer to the generality problem. It is notable, as an aside, that were the PIF not taken to offer us a solution to the generality problem, a counterexample to it that showed a particular population or set of populations to which it did not apply would be neither surprising nor germane. Biologists model the fitnesses of organisms in specific kinds of populations (with a particular genetic system, population size, and so on) in different ways throughout the biological literature. But given that the PIF does claim to offer a solution to the generality problem, these putative counterexamples are taken to constitute a significant problem for the PIF.

Before turning to these counterexamples, however, we would like to highlight one key distinction that is frequently overlooked. The distinction is that between the PIF as a philosophical position or an 'interpretation' in a broad sense—the claim that the fitness of an organism is traceable to the probability distribution over its possible numbers of offspring produced—and the mathematical model which reduces that propensity to a scalar value (which we will henceforth call a 'model of' the propensity), as expressed, for example, by Brandon's Equation (1). Importantly, all the counterexamples raised against the PIF are counterexamples to the mathematical model. They demonstrate that the fitness of an organism can change without the fitness value determined by Equation (1) changing. This point will be vital when we turn to describing the different ways in which we can respond to these counterexamples. The three most devastating such issues that have been raised are the moments problem, the delayed selection problem, and the timing of offspring problem. Let's consider them briefly in turn.

1.2.1 The moments problem

Brandon's equation computes the weighted average of O's possible offspring. But individual fitness, it turns out, is sensitive not only to the average number of possible offspring but also to higher moments of the possible offspring distribution, such as variance, skew, and so on. As Beatty and Finsen ([1989]), among others, have pointed out, if two organisms have the same average number of possible offspring, but one of them has a higher

variance in possible offspring, the higher-variance organism will be less fit, *ceteris paribus*.

In light of this realization, Brandon modified his equation, introducing a correction factor intended to compensate for the effects of the higher moments (Brandon [1990]):

$$A^*(O, E) = \sum P(Q_i^{OE})Q_i^{OE} - f(E, \sigma^2). \quad (2)$$

Brandon has the new element in the equation, $f(E, \sigma^2)$, 'denote some function of the variance in offspring number for a given type, σ^2, and of the pattern of variation' ([1990], p. 20).

This new formalization of the PIF treats it not as a single equation, but as an equation schema describing a 'family' of models: the exact nature of the propensity in a given case can only be specified once the details of the population are determined. This solution to the problem of the higher moments has two shortcomings. First, it is not obvious that, if this is the best model of the PIF, the PIF still offers a solution to the generality problem. Brandon has traded a single equation for an infinite disjunction of equations, and it is not clear that such an infinite disjunction can, for example, tell us anything about (or feature in) the general causal structure of natural selection and genetic drift. Second, Brandon's proposed solution does not achieve the desired generality: the following counterexamples emerged, which show that there is more to fitness than expected number of offspring corrected for the effects of the higher moments.

1.2.2 The delayed selection problem

The Q_i^{OE} are possible numbers of offspring. But there are many biological situations in which offspring production is not a good correlate for fitness. Consider the classic case of the grandchildless mutation found in the some species of the fruit fly *Drosophila* (Crow and Kimura [1956]). This mutation has no effect on the number of offspring produced, but it causes all of an organism's offspring to be sterile—that is, it has a major effect on grandoffspring production. This is a counterexample to any measure of fitness founded solely on offspring production. We might attempt to solve this problem simply by modifying the PIF to be based on an expected number of grandoffspring instead of offspring. This would solve the problem of the grandchildless gene, but other species have mutations that end in sterility not one or two, but dozens of generations later (Ahmed and Hodgkin [2000]). This is not so readily fixable, since even if the PIF was based on the current maximal number of generations necessary for all extant species, future evolution may increase or decrease the required number of generations. Brandon, it seems, would have to add a second correction factor to the infinite disjunction of equations,

making its suitability as a solution to the generality problem even more dubious.

1.2.3 Timing of reproduction

If two organisms have the same expected number of offspring, but one is disposed to reproduce earlier, then it will be fitter, *ceteris paribus*. This, too, is an effective counterexample to Equation (2), since timing of reproduction is a component of fitness but can change independently of expected offspring number. Brandon, again, could add a third correction factor, one for the timing of offspring production. But this will only make it a poorer solution to the generality problem.

1.3 The need for a new model

Thus, there are ample counterexamples to the mathematical models of the PIF expressed by Equations (1) and (2). Three different ways of responding to these counterexamples are obvious. We could (i) abandon the PIF and adopt another solution to the generality problem; (ii) argue that the PIF can be defended without a corresponding mathematical model; or (iii) look for a more robust, counterexample-free model. If we choose the first option, the statisticalist interpretation is the most obvious replacement for the PIF in this context. It offers a solution to the generality problem by abstracting over all causal details of the biological case at hand, and describing only the evolutionary outcomes in terms of their statistical distribution.

Although the statistical response is not without merit, it has considerable shortcomings. Defining selection and drift merely in terms of their population-level outcomes runs the risk of obscuring the distinction between selection and drift, as well as making trouble for the traditional ways in which biologists understand these differing contributors to evolutionary change (Millstein [2002]; Brandon and Ramsey [2007]; Millstein *et al.* [2009]). The relationship between the various accounts of causation on offer and evolutionary theory seems to produce *prima facie* evidence that drift and selection are in fact causal, *contra* the statisticalist position (Reisman and Forber [2005]; Millstein [2006]; Forber and Reisman [2007]; Northcott [2010]). Perhaps most worryingly, there seem to be instances where the statisticalist interpretation simply gets the empirical data wrong, claiming that selection and drift cannot be distinguished or that selection is not acting on a population, when in fact the opposite is true (Brandon and Ramsey [2007]; Millstein [2008]). These claims are all the subject of intense argument (see, for example, Lewens [2010]; Walsh [2010]), but it is worth our while to investigate ways in which a defender of the PIF could salvage the PIF's basic insights.

To understand and evaluate the other two possible responses to the counterexamples, we must begin by returning to the important distinction we made above between the PIF and the mathematical model of this propensity. The counterexamples just described serve as counterexamples specifically to Brandon's mathematical models of the PIF as expected number of offspring (possibly with a correction factor). In order, then, for them to serve as counterexamples to the PIF itself, two additional premises are required: (i) the formulation of a successful mathematical model of the PIF is required for the project to go through, and (ii) Brandon's original equation is either the only or the best possible mathematical model of the PIF. It is open to the defender of the PIF to reject either of these latent premises.

Perhaps the simplest way to reject the first premise would be to abandon the search for a mathematical model of the PIF entirely. We would then focus on the correctness of the PIF as a philosophical understanding of fitness, ignoring—or leaving to the biologists—the matter of determining the precise mathematical details of how this interpretation of fitness might be formalized. Alternatively, we could reject the second premise, and resume the search for a new mathematical foundation for the PIF. It is this latter approach, we believe, which stands the best chance of solving the generality problem while saving the possibility of a causal interpretation of fitness, natural selection, and genetic drift.

We must, however, defend this choice. Why is it that the correct response for a defender of the PIF is to continue the search for a mathematical model? Might we not best interpret many of the arguments of the statisticalists as proving to us that such a search is likely to be fruitless? We claim that it is not. First and foremost, if a counterexample-free model of the PIF can be developed, this implies the tenability of the PIF-as-interpretation. While the lack of a model does not imply the incoherence of the PIF, a counterexample-free model shows that the PIF can be formalized in a clear and explicit manner. And further, the development of models of the propensity interpretation that are connected with biological practice can form a bridge between philosophical theory and scientific practice. If we can craft a model of the PIF that connects it with current work in biology, then the PIF—which otherwise may seem esoteric and non-biological—can be shown to be directly tied to contemporary evolutionary theory.

We argue that the lack of recognition of the complexity of the biological world has been one of the key mistakes made by defenders of the PIF. Beginning with expected number of offspring one generation into the future and then adding a host of correction factors has been, as we have seen, an ultimately fruitless path. We will instead discard this formulation and begin afresh, offering a model that does indeed avoid these counterexamples, and connects directly with the much more sophisticated mathematical models

arising from cutting-edge mathematical biology. Our model can thus serve as the new foundation for the PIF.

2 A New Formalization

Let us begin by carefully considering the structure of the PIF itself—that which any mathematical model of it is intending to capture. (Henceforth, when we refer to 'the PIF', we intend to denote the PIF-as-philosophical-interpretation, not any particular model thereof.) We will begin with the sketch offered by Ramsey's 'block fitness' ([2006]), attempting to provide it with some mathematical rigour. This formal structure, at this point in the argument, is not intended to capture any specific mathematical model of the PIF. That is, we are not yet offering our own model of the PIF. Rather, the following is meant to give us a vocabulary in which any mathematical model of the PIF might be phrased (including, as we will see later, Brandon's original Equation (1)). The formalism offered here will be quite general, containing far more terms than are required by most models of the PIF. This would allow one to formalize models of any complexity in these terms by simply ignoring terms for elements not required for the model.

Consider an individual organism, o, in a given environment, E, with a given genome, G.[3] Over time, o may produce a population of offspring, o_i^1. And these, in turn, may produce offspring o_i^2. We say may produce because, for each organism (in each generation), there is a set of possible reproductive outcomes for that organism's life. These ways that organisms might live (or 'possible lives' for short[4]) might include dying early of malnutrition, being preyed upon as a juvenile, or living to maturity and producing many offspring. These possible lives reflect not just the overall reproductive output of an organism (as used in Brandon's Equation (1)), but many other features besides, such as the timing of offspring production. It is thus clear that mathematical models of the PIF can, if we wish, rely on many more theoretical resources than are utilized by standard formalizations such as Equations (1) and (2).

Let us return to the further development of our theoretical vocabulary. Combining these possible lives over generational time leads to a set of 'possible daughter populations' of o.[5] There are many such possible sets, each containing all the descendants that o might produce in some set of

[3] We do not intend a particularly restrictive definition of either 'environment' or 'genome' here. Genome, for example, should be taken to include all heritable factors passed on from parent to offspring, including epigenetic and behavioural transmission.

[4] No particular modal ontology should be read into these 'possible lives'. In fact, we believe—consistent with the propensity interpretation—that these possible lives are best understood as the manifestation of a probabilistic dispositional property over time, a propensity.

[5] We should note here that these are not 'populations' in any sense familiar from population genetics or ecology. A more appropriate term might be 'lineage', but we wish to avoid confusion with several current theories of 'lineage fitness' (see below).

circumstances. Call each of these possible daughter populations ω_i, and call the totality of such possible daughter populations (the set containing all of them) our sample space Ω.[6]

Now that we have our set of possible daughter populations, we need a way of tracking how probable these various possibilities are. To do this, we define a σ-algebra and probability measure, $\mathcal{F}$ and Pr, over Ω, in the traditional way. The details of this operation need not concern us here; this is the traditional mathematical formalization that lets us assign probabilities to the elements of our sample space. The probability $Pr(\omega_i)$ assigned to each possible daughter population is simply the probability that ω_i will be the actual daughter population of o.[7]

This set of possible daughter populations along with their associated probabilities clearly gives a very thorough picture of the 'success' of o. But we have yet to offer a mathematical model of the PIF—merely a very precise, perhaps unnecessarily large, mathematical vocabulary in which many various mathematical models of the PIF might be phrased. These raw sets of possibilities and their associated probabilities cannot, for example, be directly compared to produce a fitness rank ordering. The task of constructing our novel mathematical model of the PIF from these elements, then, is the aim of the remainder of this section. We are searching for the measure of individual fitness that can be extracted from this expansive set of theoretical resources which is maximally sensitive to the features of the raw sets of possibilities, and therefore as free as possible from the sorts of counterexamples articulated in Section 1.2.

As we discussed in Section 1, the traditional way of turning the raw sets of possibilities and their associated probabilities into scalar values, expressed in the original formulation of the propensity interpretation (Brandon [1978]; Mills and Beatty [1979]), is the following: First, define a function $\phi(\omega, t)$ which computes the size of some particular possible daughter population, ω, at time t. Then, fixing T as the time one generation into the future, we define individual fitness as

$$F = \int_{\omega \in \Omega} Pr(\omega) \cdot \phi(\omega, T) \mathrm{d}\omega. \tag{3}$$

This is simply the continuous analogue of the weighted average. We multiply the probability of each possible life by the size of the daughter population for that outcome one generation in the future, and then integrate to obtain the expected daughter-population size in the next generation. This is precisely

[6] Equivalently, one could define Ω as the space of functions $\omega : \mathbb{R} \to \mathbb{R}^n$, which take a time $t \in [0, \infty)$ to a 'state vector' consisting of some finite number of real-valued degrees of freedom. The 'possible daughter population' formulation, however, is more perspicuous, so we will use it for the remainder of the discussion.

[7] Note that the sample space Ω is uncountable, necessitating that we integrate with respect to the probability measure Pr. See Appendix A for information on the size of Ω, as well as a proof that Ω is well-behaved.

equivalent to the result obtained by the traditional PIF, giving the same numerical results as Equation (1). That is, this is just Brandon's original formulation of the PIF expressed in our new theoretical vocabulary.

As mentioned, however, this formulation is subject to many problems. First and foremost, we need to remove the reliance of Equation (3) on T, and hence on the daughter-population size only one generation into the future, to resolve the delayed selection problem. A first attempt at removing this time-dependence might lead us to compute something like the limit $F_\infty = \lim_{T\to\infty} F$, computing individual fitness in the 'infinite long run'. This would assuredly accomplish our goal of capturing all causal influences which might impact the future fate of an organism within a population.

There is, however, no guarantee that this infinite limit converges, is finite, connects with biological practice, or in any way tracks other measures of individual fitness. First and foremost, it seems quite likely that, in all cases, $F_\infty = 0$. If every possible daughter population of o goes extinct in the infinite long run, then for every $\omega \in \Omega$, $\lim_{t\to\infty} \phi(\omega, t) = 0$, and thus $F_\infty = 0$. On the other hand, if some possible daughter populations do not go extinct, then it is possible that the population dynamics at infinite time are so chaotic that the limit in F_∞ does not converge to a stable value.

How, then, can we produce a long-run measure of individual fitness from Ω? As it turns out, the problem of determining this function is equivalent to a well-studied issue in demographics and mathematical biology: the asymptotic behaviour of sequences of random, non-negative matrices. Results in this theoretical arena (following Tuljapurkar and Orzack [1980]; Caswell [1989]; Tuljapurkar [1989], [1990]) can guarantee the existence of a limit much like F_∞. (The details of this derivation may be found in Appendix A.) These results allow us to define individual fitness instead as

$$F = \exp\left(\lim_{t\to\infty} \frac{1}{t} \int_{\omega\in\Omega} Pr(\omega) \cdot \ln(\phi(\omega, t))\mathrm{d}\omega\right). \tag{4}$$

Before we consider the peculiarities of this new model (such as its infinite limit, logarithms/exponents, and factor of $1/t$), let's examine how it solves the problems of extinction and chaotic future dynamics. We must enforce three assumptions on the possible daughter populations and their associated probabilities to guarantee that the limit in Equation (4) converges (Tuljapurkar [1990], p. 25): (i) demographic weak ergodicity; (ii) that a random and stationary process generates the Pr function; and (iii) that the logarithmic moment of the growth rate is bounded. For the sake of brevity, we will pass over the detailed mathematical characterization of these assumptions here.[8] In short,

[8] For demographic weak ergodicity, the reader may consult (Seneta [1981], pp. 80–91; Tuljapurkar [1990], p. 17; Tuljapurkar and Orzack [1980], pp. 319–20; Cohen and Heyde

demographic weak ergodicity assures that there exists some non-zero probability of the population's survival at all times, t, getting us around the extinction problem mentioned above. While this assumption is biologically unrealistic, it is quite common in mathematical demography, and can be dealt with either by describing extinction as a threshold (that is, a population 'goes extinct' when its size falls to less than some small value, n), or by introducing some random environmental sampling variation into the model (see Keiding [1975]). A guarantee that the *Pr* function is generated by a stationary random process assures that chaotic population dynamics are not permitted. (The boundedness of the logarithmic moment of the growth rate is of purely technical interest.) In general, however, we defer here to Cohen, who states that 'under reasonable conditions, which are likely to be satisfied in demographic applications, the stochastic process and the matrices in Leslie form [the population analogue of our individual daughter populations and probabilities] are such that the limits in [Equation (4)] exist' (Cohen [1979], p. 164).

These three conditions do, however, have biological significance for our model. Most significantly, they imply that the selective pressures at work are density-independent, and that the population dynamics are non-chaotic. These are the most substantial limitations of our model and, because of this, Equation (4) clearly cannot provide the PIF with a complete answer to the generality problem. However, the necessary mathematical work to generalize our derivation to cases of non-static environments and density dependence, as well as chaotic population dynamics, has been published within a research program known as adaptive dynamics, to which we will return shortly. We omit it to simplify our derivation, as it relies on a hefty theoretical apparatus which considerations of length and accessibility prevent us from presenting here.[9] Our Equation (4) is the density-independent, non-chaotic limit of this more sophisticated work, and thus, given these restrictions, is equivalent to this more general model. Further, and perhaps most importantly, all the counterexamples that have been offered in the philosophical literature to the traditional mathematical model of the PIF (Equation (3)) are resolved by Equation (4).

Let us take stock. We have explicated the PIF itself via a very extensive picture of the success of an organism, o, by considering all the possible daughter populations to which it might give rise, and we have then modeled this

[1985], p. 123). For a discussion of the stationarity and ergodicity requirements on the random process, consult Tuljapurkar and Orzack ([1980]).

[9] The interested reader is referred to Rand *et al.* ([1994]), Grant ([1997]), Caswell *et al.* ([2004]), Caswell and Takada ([2004]) and Caswell ([2009]). Many of the most important conclusions for individual fitness follow directly from Theorem 1 of Rand *et al.* ([1994], p. 271). See Benton and Grant ([2000]) for a comparison of various approaches to fitness in population genetics and adaptive dynamics.

propensity by defining a function, Equation (4), that successfully encapsulates the behaviour of this sample space in the infinite limit.

But what does this function actually represent? Does it correspond to any other known models of individual fitness and, if so, in what circumstances? And what should we make of its peculiar mathematical features?

2.1 The new model and biological theory

Although Equation (4) was derived via a reflection on the philosophical thesis of the PIF, we will now show that the same mathematical formula has been independently derived within the field of mathematical biology. In fact, in adaptive dynamics, a variation of this equation is argued to be the optimal predictor of the fates of populations. We therefore suggest (and will attempt to demonstrate in the remainder of this section) that Equation (4) is deeply connected to biological theory and practice, and that with certain simplifying assumptions, one can derive from Equation (4) many standard fitness measures. Thus, despite the fact that it may seem less connected to biological theory and practice than the standard formulation of the PIF articulated by Brandon ([1978]), Equation (4) is much more closely connected with contemporary biological theory and practice, and is sensitive to the advances in mathematical work that have occurred in the decades since Brandon first published his attempt at providing a model of the PIF.

To explore the relationship between Equation (4) and other (recently developed as well as traditional) models of fitness, let's begin with two different models that are precisely equal to Equation (4). First, Equation (4) is equal to a model of fitness known in mathematical biology as Tuljapurkar's a, replacing the population-level quantities in Tuljapurkar's original function with Equation (4)'s individual-level quantities. Indeed, many of his results have been crucial to the derivation of our model here (Tuljapurkar and Orzack [1980]; Tuljapurkar [1989], [1990]).[10] Tuljapurkar's intent in creating this model of fitness was to produce an expanded notion of lifetime reproductive success (LRS) which can apply to the general case of environments that vary stochastically over time. Notably, Tuljapurkar's a has proven to be a successful measure of fitness in the wild. Cohen and Heyde ([1985]), for example, use it to study the growth of breeding populations of striped bass in the Potomac River. They determined that it provides a highly accurate determination of average growth rate (while noisy data make estimation of other parameters like variance more difficult).

This model has also appeared as one of the fundamental quantities in the research programme known as adaptive dynamics, which we briefly

[10] Technically, $\ln(F) = a$; see Appendix A for details.

mentioned above. Adaptive dynamics endeavours to produce a highly general notion of fitness applicable in many ecological contexts, based on two considerations: (i) the modelling of populations in variable environments, and (ii) the determination of fitness on the basis of invasion, consistent with much work throughout evolutionary ecology. In one of the seminal articles of adaptive dynamics, titled 'How Should We Define "Fitness" for General Ecological Scenarios?', Metz *et al.* note that 'the long-run growth rate' of a population in their framework 'can be defined as the limiting value, as (time) T approaches infinity, of the quantity $T^{-1}\{\ln|\mathbf{N}(T)| - \ln|\mathbf{N}(0)|\}$' ([1992], p. 198). This quantity, again, is precisely equivalent to our model.[11] As we discussed earlier, work in this field has also produced substantially more sophisticated models which can be used to take account of both chaotic population dynamics and density-dependent selection. Equation (4) is the density-independent, non-chaotic limit of these more complex models.[12]

To connect Equation (4) to further biological models, we must introduce some simplifying assumptions. First, if we assume that multigenerational effects are absent from the population, we may approximate Equation (4) by taking its value at time T, one generation into the future. Assuming that $T \approx 1$,[13] we then remove the limit and factor of $(1/t)$ from Equation (4), resulting in

$$F \approx \exp\left(\int_{\omega\in\Omega} Pr(\omega)\cdot\ln(\phi(\omega,T))\mathrm{d}\omega\right). \tag{5}$$

To further simplify Equation (5), we should note that it has the form of a geometric mean. That is, the geometric mean of a function $f(x)$ applied to some sample space X is defined as

$$GM(f(x),X) = \exp\left(\int_{x\in X} Pr(x)\cdot\ln(f(x))\mathrm{d}x\right),$$

and Equation (5) thus states that $F \approx GM(\phi(\omega,T),\Omega)$.

Notably, the use of a geometric mean in fitness models in biology is by no means a new concept (see Lewontin and Cohen [1969]; Boyce and Perrins [1987]; Sober [2001]; Simons [2002]; Lee and Doughty [2003]). In particular, however, given the prominence of this work in the philosophical community, geometric mean fitness is discussed extensively by Gillespie ([1977]), in his

[11] We omit the derivation here, as it is almost precisely identical—although phrased in a different theoretical vocabulary—to the derivation of Tuljapurkar's a presented in Appendix A.

[12] Therefore, with arguments very similar to the ones provided here, these more complex models also reduce to our model, Brandon's model of the PIF, and the other standard biological fitness measures we discuss below.

[13] We also must assume that the function $\int_{\omega\in\Omega} Pr(\omega)\cdot\ln(\phi(\omega,t))\mathrm{d}\omega$ is continuous at T, so that $\lim_{t\to T}(\int_{\omega\in\Omega} Pr(\omega)\cdot\ln(\phi(\omega,t))\mathrm{d}\omega) = \int_{\omega\in\Omega} Pr(\omega)\cdot\ln(\phi(\omega,T))\mathrm{d}\omega$. This should be true in effectively all circumstances, particularly since we have already guaranteed non-chaotic population dynamics.

summary of his earlier work. Specific translation of his work into our arena is difficult, as he is discussing the fitness of traits, rather than the fitness of individual organisms (see below for more discussion of this point). However, he notes that, in general, when we have stochasticity resulting from 'temporal fluctuations in the environment, for example, the best measure of fitness turns out to be the geometric mean of the offspring number, averaged over time' (Gillespie [1977], p. 1011).[14] He then establishes several results using the series expansion of the geometric mean, to which we will now turn.

We know from standard results in statistics that the geometric mean can be expressed as

$$\ln(GM(X)) = \ln(\bar{X}) - \frac{1}{2\bar{X}^2} M_2(\bar{X}) + \frac{1}{3\bar{X}^3} M_3(\bar{X}) - \cdots,$$

where $\bar{X}$ is the arithmetic mean of the distribution X, and $M_i(\bar{X})$ is the ith central moment of the distribution X (i.e. its variance, skewness, kurtosis, etc., for $i = 2, 3, 4, \ldots$) (Jean and Helms [1983]). Thus, we can see that when the higher moments of a distribution are negligible (i.e. when effects of variance, skew, and so on can be neglected), $\mathrm{GM}(X) \approx \bar{X}$. In these cases, we may thus consider individual fitness as though it were defined in terms of an arithmetic mean. We can therefore approximate Equation (5) by an arithmetic mean, which gives us

$$F \approx \int_{\omega \in \Omega} \Pr(\omega) \cdot \phi(\omega, T) \mathrm{d}\omega. \tag{6}$$

Equation (6), then, is just Equation (3): We have reduced our new formulation to that of the traditional model of the PIF with the aid of two relatively plausible simplifying assumptions. (Of course, in the presence of the conditions described in the various counterexamples to Equation (3) described above, these simplifying assumptions do not hold.)

Several short-term measures of fitness are particularly common in the biological literature, as they are easy to estimate and can be derived from readily available empirical data. The first is the net reproductive rate (or ratio), R_0, which is a common measure of the single-generation reproductive output of a population (Murray and Gårding [1984]; Murray [1990]). LRS is the individual analogue of this population measure. It is well-known from the literature on the original model of the PIF that Equation (6) is equal to the LRS, and thus Equation (4) reduces to the LRS. Another quite common biological fitness measure is the Malthusian parameter, the growth rate of a population given an exponential growth model. Given the LRS, we may derive the

[14] Gillespie probably has something like predictive accuracy in mind when he invokes the 'best measure of fitness', but these details need not concern us here.

(individual analogue of the) Malthusian parameter as well: $r = T^{-1} \ln(F)$ (Charlesworth [1970]; Denniston [1978]; Charlesworth [1980]; Murray [1990]).

To recap, then, our new model of the PIF as described by Equation (4) is precisely equivalent to several advanced models of fitness and, with two plausible simplifying assumptions—(i) that multigenerational effects are absent, and (ii) that effects of the higher moments are negligible—can be reduced to the original model of the PIF as well as to the most commonly used biological fitness measures (R_0, LRS, and the Malthusian parameter). Our model thus dovetails very tightly with contemporary work on the measurement of fitness in biology.

3 Possible Objections to F

Now that we have seen that Equation (4) is closely connected with biological theory and practice, let's pause to consider some of the ways philosophers might object to the model. We will, in the following section, then turn to the question of whether it successfully dodges the counterexamples discussed in Section 1.2. Our formulation is clearly a multigenerational or long-run measure of fitness. And this long-run measure of fitness makes extensive reference to not merely the organism itself, but to the organism and all its possible daughter populations—that is, our definition of fitness depends crucially on the organism's lineage. Both long-run and lineage fitness models have been challenged in the past, and in this section we will show that our model is not undermined by these challenges.

There is nothing novel about the concept of long-run fitness. In the philosophical literature, the varying time scales required in definitions of fitness have been discussed extensively by Ramsey ([2006]) and Abrams ([2009a]). In the biological literature, two prior models of very-long-term fitness have been proposed: Thoday's definition of fitness as 'the probability that [. . .] a unit of evolution will survive for a given long period of time, such as 10^8 years' ([1953], p. 98), and Cooper's definition of fitness as 'expected time to extinction' ([1984]).

Our new model is clearly not equivalent to either of these models and, we will argue, neither of them can serve as a satisfactory answer to the generality problem in the context of the PIF. To see why, let us first consider Thoday ([1953]). Thoday's model simply sets a large future time (10^8 years) and calculates the probability of the survival of a given organism's lineage to that point. The primary objection to a model like Thoday's is simply that this time-frame is entirely arbitrary. Why should daughter-population events at time $10^8 - 1$ years be included in an organism's fitness, but events at time 10^8+1 years be ignored? It seems that no philosophically defensible answer to this question can be found. It is for this reason that we have attempted to include all possible future causal influences in the scope of our model. Further,

should evolutionary dynamics be chaotic (a possibility introduced in the last section), there is no guarantee that the probability to which Thoday refers will even have a definite value.

Cooper's ([1984]) definition of fitness as expected time to extinction suffers from a different, but equally fatal flaw. The precise numerical expected time to extinction of a given organism seems to have only a very loose relationship to other more commonly used models of fitness. Expected time to extinction may well be correlated with individual fitness, and one could obviously derive expected time to extinction from the theoretical resources offered by our set of possible daughter populations. But for expected time to extinction to serve as a fundamental model of fitness, this derivation would need to be reversible—one would need to be able to derive other standard models of fitness (such as those mentioned at the end of the last section) *from* expected time to extinction, which is impossible. Expected time to extinction thus solves the generality problem at the price of alienating the PIF from all other work on the concept of fitness in both philosophy and biology. Given that our model does not suffer from this flaw, we believe this price is too high.

Our model, despite being long term, is multigenerational in a different manner than either of the models of Thoday and Cooper. Its time scale is infinite, encompassing all possible future causal influences on organisms. At the same time, it still reduces to other common short-term models of fitness. It thus effectively avoids the problems that Thoday and Cooper fall prey to.

Finally, we will consider several other objections that have been raised against both long-run and individual models of fitness.

3.1 Objection 1: Natural selection is short term

Brandon argues, referring to Thoday and Cooper's long-term notions of fitness, that 'they fail to explain how the process of natural selection can be sensitive to differences in long-term probabilities of surviving offspring'. Selection, he notes, is

> the diferential reproduction of phenotypes that is due to the differential adaptedness of those phenotypes to a common environment. Evolution results from this process if the phenotypic differences are heritable. How could this process be sensitive to long-term probability (i.e. over many generations) of surviving offspring? (Brandon [1990], p. 25)

Long-term notions of fitness are, that is, irrelevant to the process of natural selection—since selection is short term, individual fitness must be short term as well.

In response to this charge, Sober notes, quite correctly, that 'the fact that selection occurs one generation at a time does not mean that it is wrong to define a quantity that describes a trait's long-term expected fate' (Sober [2001],

p. 313).[15] To consider a similar example, just because the half-life of a particular sample of uranium may be defined in terms of its expected behaviour thousands of years into the future, we need not say that this somehow means that radioactive decay does not 'act on' the uranium 'as it currently is'. Radioactive decay has no more foresight than natural selection—and both may make reference to future (or even possible future[16]) events.

One might reply that the behaviour of radioactive decay is somehow more 'regular' or 'predictable' than the behaviour of biological organisms, and that therefore this analogy fails.[17] But this isn't the relevant feature of the analogy (if, in fact, the behaviour of uranium is any more 'regular' than that of organisms, which is itself not obvious). The objection as argued by Brandon seems to claim that it is a category mistake to include future events in the definition of fitness, as selection acts only in the present. A half-life certainly does make reference to future events for its definition, and radioactive decay certainly acts only in the present. This facet of the analogy thus clearly holds. This objection does not, therefore, challenge the coherence of our model.

More importantly, however, this objection misunderstands the purpose of our model. We are attempting to produce a model of the PIF that forms a successful solution to the generality problem. Solutions to the generality problem require casting an expansive net, including the extension of our model of fitness to the long term. Other models of fitness will of course be useful in other pragmatic contexts—the generality problem is often (or even usually) far from our minds when we work on evolutionary systems. And our demonstration in the last section that our model reduces, in the short term (and given several other limiting assumptions), to several well-known biological models of fitness can give us hope that this new model of the PIF can both provide a solution to the generality problem and ground a theoretically unified picture of fitness.

3.2 Objection 2: Descendants are only minimally related to ancestors

Another objection to long-term fitness is offered by Ramsey. He writes that since, over time (for sexually reproducing species), the organisms constituting the daughter population of some organism are related less and less to that organism (a factor of $\frac{1}{2}$ in the first generation, $\frac{1}{4}$ in the next, and so on), the long-term descendants of two organisms in a population may well be roughly

[15] Similarly Abrams ([2009b], p. 751) argues that 'since probabilities of long-term effects can be derived from probabilities of short-term effects, the former are simply mathematical properties of causes acting in the short term'.

[16] If there had been only one molecule of uranium in the entire universe, that molecule would still have the same half-life as the uranium which we know, even though this half-life could be defined in terms of other possible (not actually existing) atoms of uranium.

[17] We thank an anonymous reviewer for pointing out this possible reply.

identically related to each of those ancestors. It is therefore a mistake to think that the fitness of those two organisms somehow depends on the characteristics of those descendants.

Two responses can be made to this objection. First, Ramsey is at this point concerned with the operationalization of his notion of fitness. It is true that as we move from the short term to the long term, we begin to consider organisms that may be quite different than the original organism we intend to study. And it also may be true that the precise fitness difference between two organisms hinges upon an organism that we cannot measure, for it lies many generations into the future. But we have already seen that in many relevant cases, the long-term notion of fitness we describe here reduces to easily measurable short-term fitness measures. Of course, these require simplifying assumptions, and will therefore occasionally produce the wrong answers. In this case as in many others, we must assess the common biological tradeoff between complex, accurate biological models that are difficult to measure and simpler, less accurate models that are more empirically tractable.

Second, another way of reading this objection would confuse an accurate point about trait fitness with an inaccurate claim about individual fitness. When we are considering the fitness of traits, Ariew and Lewontin (among others) remind us that 'the rate of reproduction *by* a genetic type is not the same as the rate of reproduction *of* a genetic type' ([2004], p. 352). In measuring the long-term fitness of a trait, then, the genetic relatedness of those offspring with their ancestors is, in fact, a crucial point. If we lose track of this relatedness, we run the risk of inflating the fitness of ancestral types (a mistake which Ariew and Lewontin ascribe to Fisher).

In the case of individual fitness, on the other hand, such a worry does not arise. Individual fitness, as we have explicated the picture provided by the propensity interpretation, is concerned with the size of the daughter population of a given individual. Membership in a given daughter population does not come in degrees—one is either a descendant of a given organism or one is not. If a given future organism is a descendant, then it is counted by our model.[18] While the 'coefficient of relatedness' of some arbitrary, distant descendant to two members of the ancestral population may be both similar and very small, this gives us no reason to think that the fitnesses of these two ancestors will be equivalently similar. Effects of path dependence on the daughter populations in the intervening generations, for example, may well be quite significant.

[18] Of course, there are manifold issues concerning individuality in evolutionary biology, and it thus may not always clear when an 'organism' is to count as a 'descendant' (Godfrey-Smith [2009]). Our model, however, suffers from this problem no more acutely than any other of the many theories in evolutionary biology that require the counting of individuals.

3.3 Objection 3: Evolutionary time scale is pragmatically determined

Sober claims that we should wish to remain agnostic over whether we should choose long-term or short-term measures of fitness. 'Long-term fitness', he writes, 'is a coherent concept that may be useful in the context of certain problems; however, its coherence and desirability do not undermine the concept of short-term fitness' (Sober [2001], p. 313). In general, we agree. As noted above, we have introduced our model with the intent of cementing the place of the PIF (and, along with it, the causal interpretation of fitness, selection, and drift) as one possible solution to the generality problem. But short-term predictions, as we have seen above in our discussion of the relationship of this model to other biological definitions of fitness, can readily be derived from this model under various plausible sets of limiting assumptions. Thus, we are fully entitled to utilize these short-term predictions when they are needed. And as we have already noted, the solution to the generality problem—an active issue in the philosophy of biology—requires the model to take the long-term view.

3.4 Objection 4: Long-term fitness is lineage fitness

We claim to be setting out a model of individual or organismic fitness. Our derivation, however, seems to traffic only in lineages or daughter populations. Isn't this model, then, in fact a model not of individual but of lineage fitness?

We should begin by noting that we do not intend to disparage the usefulness of lineage fitness. Jost, who defines lineage fitness as 'the number of descendants weighted by their degree of relatedness with the ancestor of the lineage' ([2003], p. 331), has made a persuasive argument for lineage fitness and used it effectively to analyze the emergence of altruism (although see Okasha ([2006]) for a criticism of this concept).

It is true, however, that the objection that Equation (4) represents lineage fitness instead of organismic fitness would defeat our claim to be offering an improved model of the PIF. But reading our model as a model of lineage fitness misunderstands our work. Clearly, the possible future lineages of an organism are one of the determinants of its fitness. But this is not equivalent to the claim that lineages are the bearers of fitness. Our model, that is, is a model of individual fitness that depends on lineage characteristics.

3.5 Objection 5: The theory of evolution by natural selection fundamentally concerns trait fitness, not individual fitness

Our model is, as we have stated, a model of the fitness of individual organisms. Organismic fitness plays important roles in parts of ecology and evolutionary

biology, and is the concept of fitness underlying the PIF. On the other hand, much of the active work of biologists, as well as many of the arguments of philosophers, relies instead upon trait fitness. Furthermore, many counterexamples have been raised against propensity-based models of trait fitness, including critiques by Ariew and Lewontin ([2004]),[19] and Krimbas ([2004]).[20] One might be concerned, then, both that our model fails to respond adequately to these other counterexamples present in the fitness literature, and that our model fails to offer an account of the fitness of traits.

Trait fitness, however, is commonly understood in two different ways. First, we have trait fitness as the average fitness of all individual organisms that bear a given trait (Sober [2001]; Walsh *et al.* [2002]; Walsh [2003]). Second, we have trait fitness as a prediction of future trait prevalence: the quantity that lets us predict the frequency of a trait in the next generation given its current frequency (Matthen and Ariew [2002]; Walsh [2003]). If the first of these two definitions is adopted, then trait fitness is straightforwardly parasitic on individual fitness, and a model of individual fitness must be provided to make sense of the fitness of traits. If the second definition is adopted, however, then we are dealing with quite a different quantity than the one modelled here. Trait fitness in this second sense relies on individual fitness as well, but also includes factors such as heritability. Thus, under either of the standard ways of understanding trait fitness, individual fitness is in some sense foundational. Trait fitness values are either directly derived from individual fitness values, or individual fitness values are a component of trait fitness. Because of this, we are justified in simply providing a model of individual fitness as the foundational concept in the PIF.

4 Response to Counterexamples

In the first section, we examined three counterexamples to the traditional model of the PIF as expected number of offspring. In this section, we will show that our new model of the PIF, as represented by Equation (4), does not fall prey to these counterexamples.

4.1 Timing of reproduction

If two organisms, O_1 and O_2, have the same expected number of offspring, but O_1 is disposed to reproduce earlier than O_2, then it will be fitter, *ceteris paribus*. The reason for this is that O_1 (and presumably its offspring, if the

[19] Ariew and Lewontin intend their critique to be targeted at 'a scalar quantity [. . .] which then predicts changes in the representation of types' ([2004], p. 350).

[20] Although he speaks occasionally of the fitness of individuals, Krimbas's main concern is 'the absolute or Darwinian fitness of a certain genetic constitution of individuals of the same species in a population' ([2004], p. 190), clearly a notion of trait fitness.

trait responsible for its different reproductive behaviour is heritable) will have a shorter generation time. If both O_1 and O_2 have two offspring, but O_1 has them twice as early in life, then O_1 will have twice as many descendants as O_2 when O_2 finishes reproducing. If this is not a dramatic fitness difference, then nothing is. Yet, if we measure the fitness of O_1 and O_2 with Equation (2), we arrive at the same value. Thus, even the revised version of the original PIF does not take this fitness difference into account. The reason for this is that the equation is a function of possible lifetime reproductive success (LRS). That is, it merely tallies the reproductive event outcomes of entire life histories without being sensitive to other important properties of these life histories, such as the temporal arrangement of reproductive events.

Equation (4), however, solves this problem. It might seem that the reason that it does so is that the model is multigenerational. That it extends through a large number of generations in determining fitness values. Although the multigenerational nature of our model is important, what is crucial for solving the timing of reproduction problem is that our model uses time instead of number of generations to determine fitness. If the model was multigenerational but used generations instead of time to determine fitness, then O_1 and O_2 would be deemed equally fit, since they would both have the same number of grandoffspring, great-grandoffspring, and so on. But using time instead of generations leads to a higher fitness value for O_1 than O_2. It is important, however, that the right time scale is used. If O_1 reproduces at the age of five and O_2 at ten, then defining fitness as the expected number of offspring at one year would mistakenly compute both of their fitness values to be zero. The time scale has to be at least as large as the longest generation time for any individual whose fitness is being compared, but it will regularly need to go far beyond a single generation. The fact that Equation (4) is based on an infinite limit clearly provides us with a sufficiently large time scale to account for all possible variability in reproductive timing.

4.2 Delayed selection

Not all offspring are created equal. Some will share the fertility of their parents, whereas others will be infertile. The grandchildless (and other, longer-term) mutations mentioned in the first section, which have been found in *Drosophila* and other taxa, show that a single-generation time scale is not sufficient to capture fitness. One might, however, think that the grandchildless mutation does not necessitate a super-generational time scale. Instead, one could attempt to preserve single-generational fitness by simply discounting offspring based on their fertility—infertile offspring would not be counted as offspring. This proposal, we argue, merely smuggles in a multigenerational time scale. Consider, for example, why it is that we must not count

O's infertile offspring. The correct answer to this, it seems, is that they will not give O any grandoffspring. Thus, the choice to discount the fitness of infertile offspring relies on fitness being multigenerational. Furthermore, both fertility and viability affect fitness, and both come in degrees. Thus, not only would infertile individuals have to be excluded, each offspring would have to be assigned a weight, depending on its relative fertility and viability. But what is the assignment of such a weight, but an assignment of fitness values to the offspring? There are thus two reasons why the offspring-weighting proposal fails. First, it defines fitness in terms of fitness. Second, it defines the fitness of the offspring in terms of the grandoffspring, and so on, making it in fact multigenerational.

Fortunately, Equation (4) is sensitive to these and similar mutations. Determining the fitness of individuals via their descendant pool 'at infinity' allows us to incorporate the fact that not all descendants are created equal. Our fitness model takes account of the unlimited variability of fertility and viability, as the long-term descendant pool is sensitive to these factors.

4.3 The moments problem

As we saw in Section 1, fitness is a function of not only the first moment (i.e. arithmetic mean) of the distribution of possible offspring, but the higher moments as well (variance, skew, and so on). This problem was not recognized in the original formulation of the PIF, but was later recognized and solutions were offered. The solution offered by Brandon ([1990]), as we saw (Equation (2)), was to add a correction factor that would discount the fitness of individuals based on the structure of the distribution of possible offspring. But this solution was unsatisfying, since it merely offered 'some' unspecified function of this distribution.

Our model, unlike previous models, solves the moments problem through the introduction of a concrete, specific function, Equation (4). This function solves the moments problem by virtue of its long time scale. To see why this is the case, consider a simple example of two organisms, O_1 and O_2, which reproduce asexually and clonally, and have discrete generations. Each generation, the first organism and its descendants have $Pr(\phi(\omega, T) = 1) = 1$ (that is, the probability of having 1 offspring each generation is 1), and the second organism and its descendants have $Pr(\phi(\omega, T) = 0) = 0.5$ and $Pr(\phi(\omega, T) = 2) = 0.5$ (equal probability of having either 0 or 2 offspring). Both O_1 and O_2 have the same expected number of offspring after one generation (namely, one), but O_1 is fitter. To see why, consider what you would expect the daughter population size of O_1 to be as $t \to \infty$. Since O_1 has a probability of 1 of having 1 offspring, the daughter population size will be one. But O_2, on the other hand, will not fare so well. The probability of the

population going extinct will approach 1 as $t \to \infty$, since extinction is all but guaranteed in the long run: in each passing generation, each of O_2's descendants runs a 50% risk of lineage extinction. Taking the long view, as does Equation (4), correctly evinces the superior fitness of O_1. We lack the space to provide more examples of higher moments here (such as those presented in Beatty and Finsen ([1989]) and Abrams ([2009b])), but they reveal themselves in the long term as well.

5 Conclusion

There has long been a perception that the PIF is in dire straits. As we have seen, it has from its earliest days purported to offer us a solution to the generality problem, yet the mathematical model of it commonly offered is subject to a host of counterexamples. Many or even most real-world biological populations are subject to one of the difficulties described, making the PIF appear to be an inadequate answer to the generality problem—and thus not able to support the causal interpretation of fitness, selection, and drift against its statisticalist opposition. Propensity theorists have tried to save the PIF by amending its standard mathematical model with correction factors. These amended versions have saved the PIF from counterexamples, however, only by rendering it a poor solution to the generality problem. Furthermore, as the statisticalists have eagerly pointed out, these proposed correction factors do not seem to dodge all of the counterexamples. That is, there still has not been a model of the PIF offered in the literature robust enough to withstand the full array of philosophical difficulties with prior models.

As we argued in Section 1, however, all these counterexamples are problems not with the PIF itself, but with the various mathematical models of it which have been proposed. The opponents of the PIF have done nothing by way of arguing that the extant mathematical models of the PIF are either the only or the best ways to formalize this interpretation of fitness. We considered three possible ways for a proponent of the PIF to respond. One could (i) jump ship and embrace the statisticalist interpretation of fitness, selection, and drift (or something like it); (ii) abandon the search for a mathematical model of the PIF entirely; or (iii) craft a model of the PIF that can offer it a new mathematical foundation. In this article, we have accomplished the third option. We have constructed a model that retains the PIF's purported ability to solve the generality problem without being subject to the counterexamples that have been proposed against it.

We have not, notably, argued directly against the statisticalist position itself. As we have argued above, the statisticalist position is not without difficulty, and has been criticized on many fronts (Millstein [2002]; Brandon and Ramsey [2007]; Millstein *et al.* [2009]; Northcott [2010]; Otsuka *et al.* [2011]; Ramsey

[forthcoming]). One of the seemingly compelling arguments in its favour, however, is that it is not subject to the sorts of counterexamples that undermine the PIF. Since we have provided a model of the PIF that avoids these counterexamples (and, hence, have demonstrated that there is no reason to think that the PIF cannot solve the generality problem), one of the main reasons for adopting the statisticalist position has been called into question. The statisticalists, we suggest, would need to attempt to form counterexamples to this new model, not the older formulation provided by Brandon ([1978]).

While this by no means resolves the debates over the role of fitness in evolutionary theory—one could, even taking our model into account, still reject the notion of fitness as a causal property—we hope that the presentation of a mathematical model that resists the now-common counterexamples to the PIF will allow for a more sophisticated debate over the nature of fitness, natural selection, and genetic drift.

Appendix A: Derivation of Fitness Model

Begin with the set Ω of possible daughter populations of some individual, o, as defined above. We noted that we define a σ-algebra and probability measure, $\mathcal{F}$ and Pr, over Ω, in the usual way. To do so, however, we must demonstrate that Ω is sufficiently well-behaved that a probability measure may be defined over it—one might plausibly think that this set is simply too large to be suitable as a sample space. Consider our provisional definition (from footnote 6) of Ω as the space of functions from $\mathbb{R} \mapsto \mathbb{R}^n$. This set has cardinality $N(\Omega) = 2^{2^{\aleph_0}}$, which makes it impossible to establish a standard probability measure over Ω in the normal manner—by exhibiting an isomorphism between Ω and either $\mathbb{R}$ or $[0, 1]$.

However, as demonstrated by Nelson ([1959]), we are able to define a standard σ-algebra $\mathcal{F}$ and a Borel probability measure Pr over certain subsets of this larger set. Namely, we can get what we need if we restrict our attention to (i) continuous functions ω; (ii) functions ω with only point discontinuities; or (iii) functions ω with only jump discontinuities.

Although we would like to remain neutral on how the n-degrees of freedom available in the state vectors at each time t might be parameterized, it doesn't appear out of the question to assume that the ω functions will have at worst jump discontinuities. If this is the case, then the proofs in Nelson ([1959]) demonstrate that a working probability measure can be reasonably defined over Ω.

With the sample space suitably defined, we may now derive our model. As we did above, define $\phi(t, \omega)$ as the function that takes a particular point in the sample space ω and a time t to the number of o's progeny living at time t on that outcome.

Now return to the problem of constructing a long-term picture of organismic fitness. We cannot evaluate the ϕ function at time $t = \infty$, as we will obtain (at least potentially) infinite values. We thus need to define some function of the ϕ values which will converge as t goes to ∞. We know, however, as stipulated above, that the ϕ values are generated by a stationary random process, that demographic weak ergodicity holds, and that the logarithmic moment of vital rates is bounded. From this we may conclude (Tuljapurkar [1989], pp. 233–4) that the following limit exists:

$$a = \lim_{t\to\infty} \frac{1}{t} E_w \ln(\phi(t)), \tag{7}$$

with E_w an expectation value weighted by the probabilities given by our σ algebra as defined above, and removing the parameter ω from the ϕ function when it appears inside a mathematical expectation. If we take the exponent of both sides of Equation (7), we arrive at

$$F = \exp\left(\lim_{t\to\infty} \frac{1}{t} E_w \ln(\phi(t)) \right). \tag{8}$$

This value F is, then, the value of fitness in our model: Equation (8) is equivalent to Equation (4). It is precisely equal to the exponential of Tuljapurkar's a (that is, Equation (7) just *is* Tuljapurkar's a), and therefore roughly equal (under simplifying assumptions) to the net reproductive rate and related to the Malthusian Parameter ($r = \ln(F)/T$, with T the generation time).

Acknowledgements

We would like to thank Robert Brandon, David Crawford, John Endler, Agustín Fuentes, Alastair Grant, Douglas Futuyma, James Justus, Elliott Sober, and several anonymous referees for comments throughout the drafting of this article. The article has undergone major evolutionary transformations based on these very helpful comments, and some of the early readers will hardly recognize this article as being the same as the ancestral one that they read long ago.

Charles H. Pence
Department of Philosophy
University of Notre Dame
Notre Dame, IN 46556
charles@charlespence.net

Grant Ramsey
Department of Philosophy
University of Notre Dame
Notre Dame, IN 46556, USA
grant.ramsey@nd.edu

References

Abrams, M. [2007]: 'How Do Natural Selection and Random Drift Interact?', *Philosophy of Science*, **74**, pp. 666–79.

Abrams, M. [2009a]: 'What Determines Biological Fitness? The Problem of the Reference Environment', *Synthese*, **166**, pp. 21–40.

Abrams, M. [2009b]: 'The Unity of Fitness', *Philosophy of Science*, **76**, pp. 750–61.

Ahmed, S. and Hodgkin, J. [2000]: 'MRT-2 Checkpoint Protein is Required for Germline Immortality and Telomere Replication in *C. elegans*', *Nature*, **403**, pp. 159–64.

Ariew, A. and Ernst, Z. [2009]: 'What Fitness Can't Be', *Erkenntnis*, **71**, pp. 289–301.

Ariew, A. and Lewontin, R. C. [2004]: 'The Confusions of Fitness', *British Journal for the Philosophy of Science*, **55**, pp. 347–63.

Beatty, J. H. and Finsen, S. [1989]: 'Rethinking the Propensity Interpretation of Fitness: A Peek Inside Pandora's Box', in M. Ruse (*ed.*), *What the Philosophy of Biology Is: Essays for David Hull*, Dordrecht: Kluwer Academic Publishers, pp. 18–30.

Benton, T. G. and Grant, A. [2000]: 'Evolutionary Fitness in Ecology: Comparing Measures of Fitness in Stochastic, Density-Dependent Environments', *Evolutionary Ecology Research*, **2**, pp. 769–89.

Bouchard, F. and Rosenberg, A. [2004]: 'Fitness, Probability, and the Principles of Natural Selection', *British Journal for the Philosophy of Science*, **55**, pp. 693–712.

Boyce, M. S. and Perrins, C. M. [1987]: 'Optimizing Great Tit Clutch Size in a Fluctuating Environment', *Ecology*, **68**, pp. 142–53.

Brandon, R. N. [1978]: 'Adaptation and Evolutionary Theory', *Studies in History and Philosophy of Science*, **9**, pp. 181–206.

Brandon, R. N. [1990]: *Adaptation and Environment*, Princeton, NJ: Princeton University Press.

Brandon, R. N. and Ramsey, G. [2007]: 'What's Wrong with the Emergentist Statistical Interpretation of Natural Selection and Random Drift?', in D. L. Hull and M. Ruse (*eds*), *The Cambridge Companion to the Philosophy of Biology*, Cambridge: Cambridge University Press, pp. 66–84.

Brunnander, B. [2006]: 'What is Natural Selection?', *Biology and Philosophy*, **22**, pp. 231–46.

Caswell, H. [1989]: *Matrix Population Models*, Sunderland, MA: Sinauer Associates.

Caswell, H. [2009]: 'Sensitivity and Elasticity of Density-Dependent Population Models', *Journal of Difference Equations and Applications*, **15**, pp. 349–69.

Caswell, H. and Takada, T. [2004]: 'Elasticity Analysis of Density-Dependent Matrix Population Models: The Invasion Exponent and Its Substitutes', *Theoretical Population Biology*, **65**, pp. 401–11.

Caswell, H., Takada, T. and Hunter, C. M. [2004]: 'Sensitivity Analysis of Equilibrium in Density-Dependent Matrix Population Models', *Ecology Letters*, **7**, pp. 380–7.

Charlesworth, B. [1970]: 'Selection in Populations with Overlapping Generations, I: The Use of Malthusian Parameters in Population Genetics', *Theoretical Population Biology*, **1**, pp. 352–70.

Charlesworth, B. [1980]: *Evolution in Age-Structured Populations*, Cambridge: Cambridge University Press.

Cohen, J. E. [1979]: 'Comparative Statics and Stochastic Dynamics of Age-Structured Populations', *Theoretical Population Biology*, **16**, pp. 159–71.

Cohen, J. E. and Heyde, C. C. [1985]: 'Confidence Intervals for Demographic Projections Based on Products of Random Matrices', *Theoretical Population Biology*, **27**, pp. 120–53.

Cooper, W. S. [1984]: 'Expected Time to Extinction and the Concept of Fundamental Fitness', *Journal of Theoretical Biology*, **107**, pp. 603–29.

Crow, J. and Kimura, M. [1956]: 'Some Genetic Problems in Natural Populations', *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, **4**, pp. 1–22.

Denniston, C. [1978]: 'An Incorrect Definition of Fitness Revisited', *Annals of Human Genetics*, **42**, pp. 77–85.

Forber, P. and Reisman, K. [2007]: 'Can There Be Stochastic Evolutionary Causes?', *Philosophy of Science*, **74**, pp. 616–27.

Gillespie, J. H. [1977]: 'Natural Selection for Variances in Offspring Numbers: A New Evolutionary Principle', *American Naturalist*, **111**, pp. 1010–4.

Godfrey-Smith, P. [2009]: *Darwinian Populations and Natural Selection*, Oxford: Oxford University Press.

Grant, A. [1997]: 'Selection Pressures on Vital Rates in Density-Dependent Populations', *Proceedings of the Royal Society of London B: Biological Sciences*, **264**, pp. 303–6.

Jean, W. H. and Helms, B. P. [1983]: 'Geometric Mean Approximations', *Journal of Financial and Quantitative Analysis*, **18**, pp. 287–93.

Jost, J. [2003]: 'On the Notion of Fitness, Or: The Selfish Ancestor', *Theory in Biosciences*, **121**, pp. 331–50.

Keiding, N. [1975]: 'Extinction and Exponential Growth in Random Environments', *Theoretical Population Biology*, **8**, pp. 49–63.

Krimbas, C. B. [2004]: 'On Fitness', *Biology and Philosophy*, **19**, pp. 185–203.

Lee, M. S. Y. and Doughty, P. [2003]: 'The Geometric Meaning of Macroevolution', *Trends in Ecology & Evolution*, **18**, pp. 263–6.

Lewens, T. [2010]: 'The Natures of Selection', *British Journal for the Philosophy of Science*, **61**, pp. 313–33.

Lewontin, R. C. [1970]: 'The Units of Selection', *Annual Review of Ecology and Systematics*, **1**, pp. 1–18.

Lewontin, R. C. and Cohen, D. [1969]: 'On Population Growth in a Randomly Varying Environment', *Proceedings of the National Academy of Sciences*, **62**, pp. 1056–60.

Matthen, M. and Ariew, A. [2002]: 'Two Ways of Thinking about Fitness and Natural Selection', *Journal of Philosophy*, **99**, pp. 55–83.

Metz, J. A. J., Nisbet, R. M. and Geritz, S. A. H. [1992]: 'How Should We Define "Fitness" for General Ecological Scenarios?', *Trends in Ecology & Evolution*, **7**, pp. 198–202.

Mills, S. K. and Beatty, J. H. [1979]: 'The Propensity Interpretation of Fitness', *Philosophy of Science*, **46**, pp. 263–86.

Millstein, R. L. [2002]: 'Are Random Drift and Natural Selection Conceptually Distinct?', *Biology and Philosophy*, **17**, pp. 33–53.

Millstein, R. L. [2006]: 'Natural Selection as a Population-Level Causal Process', *British Journal for the Philosophy of Science*, **57**, pp. 627–53.

Millstein, R. L. [2008]: 'Distinguishing Drift and Selection Empirically: "The Great Snail Debate" of the 1950s', *Journal of the History of Biology*, **41**, pp. 339–67.

Millstein, R. L., Skipper, R. A. and Dietrich, M. R. [2009]: '(Mis)interpreting Mathematical Models: Drift as a Physical Process', *Philosophy and Theory in Biology*, **1**, p. e002.

Murray, B. G. [1990]: 'Population Dynamics, Genetic Change, and the Measurement of Fitness', *Oikos*, **59**, pp. 189–99.

Murray, B. G. and Gårding, L. [1984]: 'On the Meaning of Parameter x of Lotka's Discrete Equations', *Oikos*, **42**, pp. 323–6.

Nelson, E. [1959]: 'Regular Probability Measures on Function Space', *Annals of Mathematics*, **69**, pp. 630–43.

Northcott, R. [2010]: 'Walsh on Causes and Evolution', *Philosophy of Science*, **77**, pp. 457–67.

Okasha, S. [2006]: *Evolution and the Levels of Selection*, Oxford: Oxford University Press.

Otsuka, J., Turner, T., Allen, C. and Lloyd, E. A. [2011]: 'Why the Causal View of Fitness Survives', *Philosophy of Science*, **78**, pp. 209–24.

Ramsey, G. [2006]: 'Block Fitness', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **37**, pp. 484–98.

Ramsey, G. [forthcoming]: 'Organisms, Traits, and Population Subdivisions: Two Arguments against the Causal Conception of Fitness?', *British Journal for the Philosophy of Science*, doi:10.1093/bjps/axs010, Epub ahead of print, 25 January 2013.

Rand, D. A., Wilson, H. B. and McGlade, J. M. [1994]: 'Dynamics and Evolution: Evolutionarily Stable Attractors, Invasion Exponents, and Phenotype Dynamics', *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **343**, pp. 261–83.

Reisman, K. and Forber, P. [2005]: 'Manipulation and the Causes of Evolution', *Philosophy of Science*, **72**, pp. 1113–23.

Seneta, E. [1981]: *Non-negative Matrices and Markov Chains*, New York: Springer.

Simons, A. M. [2002]: 'The Continuity of Microevolution and Macroevolution', *Journal of Evolutionary Biology*, **15**, pp. 688–701.

Sober, E. [2001]: 'The Two Faces of Fitness', in R. S. Singh (*ed.*), *Thinking about Evolution: Historical, Philosophical, and Political Perspectives*, Cambridge, MA: MIT Press, pp. 309–21.

Thoday, J. M. [1953]: 'Components of Fitness', *Symposia of the Society for Experimental Biology*, **7**, pp. 96–113.

Tuljapurkar, S. [1989]: 'An Uncertain Life: Demography in Random Environments', *Theoretical Population Biology*, **35**, pp. 227–94.

Tuljapurkar, S. [1990]: *Population Dynamics in Variable Environments*, New York: Springer-Verlag.

Tuljapurkar, S. D. and Orzack, S. H. [1980]: 'Population Dynamics in Variable Environments I: Long-Run Growth Rates and Extinction', *Theoretical Population Biology*, **18**, pp. 314–42.

Walsh, D. M. [2003]: 'Fit and Diversity: Explaining Adaptive Evolution', *Philosophy of Science*, **70**, pp. 280–301.

Walsh, D. M. [2007]: 'The Pomp of Superfluous Causes: The Interpretation of Evolutionary Theory', *Philosophy of Science*, **74**, pp. 281–303.

Walsh, D. M. [2010]: 'Not a Sure Thing: Fitness, Probability, and Causation', *Philosophy of Science*, **77**, pp. 147–71.

Walsh, D. M., Lewens, T. and Ariew, A. [2002]: 'The Trials of Life: Natural Selection and Random Drift', *Philosophy of Science*, **69**, pp. 429–46.