Optimality Models and the Propensity Interpretation of Fitness

Preprint (forthcoming in Acta Biotheoretica)

Author 1: Ariel Jonathan Roffé
ORCiD: https://orcid.org/0000-0002-0051-2028
Institutional affiliations: Centro de Estudios de Filosofía e Historia de la Ciencia (CEFHIC-UNQ-CONICET), Universidad de Buenos Aires (UBA), Universidad Tres de Febrero (UNTREF)
Email: arielroffe@filo.uba.ar, ariroffe@hotmail.com

Author 2: Santiago Ginnobili
ORCiD: https://orcid.org/0000-0001-5375-965X
Institutional affiliations: Centro de Estudios de Filosofía e Historia de la Ciencia (CEFHIC-UNQ-CONICET), Universidad de Buenos Aires (UBA), Universidad Tres de Febrero (UNTREF)
Email: santi75@gmail.com

Acknowledgements: This work has been funded by the research projects PUNQ 1401/15 and SAI 827-223/19 (National University of Quilmes, Argentina), UNTREF 32/15 255 (Universidad Nacional Tres de Febrero, Argentina) and UBACyT 20020170200106BA (Universidad de Buenos Aires, Argentina).

Abstract

The propensity account of fitness intends to solve the classical tautologicity issue by identifying fitness with a disposition, the *ability* to survive and reproduce. As proponents recognized early on, this account requires *operational* independence from actual reproductive success to avoid circularity and vacuousness charges. They suggested that operational independence is achieved by measuring fitness values through optimality models. Our goal in this article is to develop this suggestion. We show that one plausible procedure by which these independent operationalizations could be thought to take place, and which is in accordance with what is said in the optimality literature, is unsound. We provide two independent lines of reasoning to show this. We then provide a sketch of a more adequate account of the role of optimality models in evolutionary contexts and draw some consequences.

Keywords: fitness, natural selection, optimality models, propensity interpretation of fitness

1. Introduction

One of the most frequent criticisms of the theory of natural selection is that its explanations are circular. Its core principle (sometimes referred to as the "principle of natural selection" or "PNS") is usually taken to have a form close to the following:[1]

(PNS) If *x* is fitter than *y*, then *x* will tend to have greater reproductive success than *y*

If one defines "*x* is fitter than *y*" as "*x* has a greater reproductive success than *y*," then natural selection would explain differences in reproductive success based on those very same differences. Given the magnitude of these objections, the issue of providing an adequate definition (or characterization) of fitness, which reflects the uses of this concept in evolutionary theory and practice, has been a central concern for the philosophy of evolutionary biology.

[1] This is only a schematic presentation of the PNS. The "tend to" here only means that fitter organisms do not always actually have a greater reproductive success than less fit organisms. This can be made more precise in different ways. For example, the "tend to" may be interpreted probabilistically, as a ceteris paribus clause, etc. There are more specific versions of the PNS in the literature (see Rosenberg and Bouchard 2015). Since the way in which one spells this out precisely is not directly relevant to our point, we stick to this more general presentation.

One of the most widely accepted definitions is the one provided by the Propensity Interpretation of Fitness (hereafter PIF), introduced with two articles by Brandon (1978) and Mills and Beatty (1979). According to the PIF, the fitness of an organism is its *ability* to survive and reproduce in a given environment.[2] That is, fitness is a *dispositional* property. Dispositional properties can typically be accessed in two different ways. Since, in the long run, they tend to manifest themselves as relative frequencies, they can be operationalized through them. For example, if a coin lands on heads 2/3 of the time in a long series of throws, then one can infer (with a certain degree of confidence) that the coin has a propensity to land on heads of about 2/3 in strength. Alternatively, one can look directly at the physical properties of the coin (such as its weight distribution) and, alongside with some physical theories or principles that connect weight distributions with landing probabilities, conclude the same thing.

Similarly, according to the PIF, fitness can be accessed via (statistically significant) extrapolation from past reproductive success, and via "direct examination" of the physical and behavioral traits of the organisms that possess them, and the way they function in the environment in which these organisms live.

As proponents of the PIF correctly argue (see, for example, Brandon and Beatty 1984), the existence of this second way of establishing fitness is important because it would allow

[2] For more precisions on this, see section 2, where we give a more complete characterization of the PIF.

biologists to explain reproductive success based on something different than (past or present) reproductive success, thus avoiding the circularity issue mentioned above. However, no one has yet presented a detailed account of how this second way of establishing fitness is supposed to work. Proponents of the PIF have suggested that this takes place through optimality models, of the kind used in behavioral ecology (Beatty 1980; Brandon and Beatty 1984; Mills and Beatty 1979; Millstein 2016). In other words, the link between optimality studies and the notion of fitness is a fundamental part of defending the genuine explanatory capacity of the PNS. Despite the importance of this point in the propensity approach, how exactly it is that optimality studies allow us to determine fitness values has not been developed yet.[3]

Our goal in this article is to expand on the topic of the connection between optimality and fitness. We show that what is perhaps the most intuitive and widespread account of the connection between these two areas is not adequate. We then suggest a more promising line of research that could be followed in order to complete this important gap in the PIF. We will also

[3] We do not claim, however, that nothing has been written on the topic, but only that existing proposals (e.g. Maynard Smith 1978; Parker and Maynard Smith 1990; see below for more references) are not sufficiently precise or well-developed to constitute a complete account, specifically for cases where the currency of the model is a proxy for fitness. See below for more on this.

show that these developments have important consequences for some accounts of the nature of evolutionary biology.

We will proceed as follows. In section 2, we introduce the PIF in more detail. Section 3 examines a specific example of an optimality model and describes a possible way of establishing a connection between these models and fitness, which is both intuitive and widespread in the optimality literature. As we will explain later on, this method consists of inferring relative fitness values from the models' relative "currency" differences. Section 4 criticizes this view. We show that inferring fitnesses from currency values, in the way proposed in section 3, requires that both currency and fitness functions correlate linearly. However, we argue that (i) linear correlation is not a requirement for an optimality model to be considered successful within behavioral ecology; and thus, (ii) linear correlation is almost never the case (or, at least, we have no reason to expect it to be). Section 5 suggests a more promising way of establishing this connection, through some empirical principles. We look at some proposals that aim to establish a qualitative (or comparative) relation between optimality and fitness. In section 6, we consider some implications of this debate for the status of the PNS and of Probability theory in evolutionary biology. Finally, we draw some conclusions.

2. The Propensity Interpretation of Fitness

There are a number of conceptual and philosophical discussions surrounding natural selection explanations and the meaning of the concept of fitness within them. Two related objections are that meaning of "fitness" renders the PNS tautological and, thus, that selective explanations are devoid of explanatory value. If one defines "*x* is fitter than *y*" as "*x* has a greater reproductive success than *y*" (as biology manuals often do, see for example Futuyma 2005, p. 272; Ridley 2004, p. 74), then the PNS claims nothing more than that organisms that have a greater reproductive success have a greater reproductive success. This claim is obviously tautological and, as such, lacks any explanatory value (for objections of this kind to natural selection theory, see Fodor and Piattelli-Palmarini 2010, p. 137; Mivart 1898, p. 272; Peters 1976; Popper 1974; Vallejo 1998).

As mentioned in the introduction, the PIF proposes that fitness should be understood as a dispositional property: the organism's *ability* to survive and reproduce.[4] More precisely, a probability distribution to leave *n* offspring is assigned to each organism. Proponents of the PIF

[4] Standard (both classical and contemporary) presentations of the PIF assume that fitness is a property of individuals. This idea has been criticized, e.g. in Sober 2013, who takes fitness to be a property of traits (he actually reaches similar conclusions to ours regarding quantitative predictions from optimality models, although parting from different starting points, and not developing them to the extent that we do, see Sober 2013, p. 338). Since our goal is to make an internal critique to this standard version of the PIF, we follow their own presentation.

usually propose some way of representing fitnesses as scalars, derived from those distributions (see Brandon 1978; Pence and Ramsey 2013, for a more recent proposal), since this is the way in which they are usually represented in population genetic models (see below).

This, in principle, avoids the issue of the tautologicity of the PNS. The ability to survive and reproduce is not conceptually the same as actual reproductive success. However, without additional precisions, there is still a risk that the PIF makes natural selection explanations unsatisfactory in a different sense. A good way to understand this is to consider some of the early objections that the account received, for instance:

> Mills and Beatty's proposal is *operationally sterile* as a definition of fitness. (…) Interpolating a disposition between fitness and actual survival and reproduction severs the direct logical connection between these three notions. But it does so only by introducing a fourth term, the disposition, which itself is unaccounted for in the theory, and so opens up again the prospects for charges of circularity it was meant to forestall. Within the theory there are no more resources for providing a non-circular explication of the propensity to reproduce than Moliere's physicians could provide for the dormitive virtue of opium. The propensity is the cause of the differences in actual rates, is transmitted from ancestors with the same propensity, and is intra-and interspecifically variable. How can we know all this about the propensity? Through its causes and effects *reflected in actual retrospective and prospective rates of reproduction*. But this is the

> circularity problem with which Mills and Beatty begin. (Rosenberg 1982, pp. 270-72, emphasis added)

The idea is that to say that *x*'s tend to have a greater reproductive success than *y*'s because they have a greater propensity to do so sounds a lot like saying that opium makes you sleep because it has a dormitive virtue. If the *only* way to know that something has a dormitive virtue is to check whether it makes you sleep (or rather, made you sleep in the past) then, again, the invocation of dormitive virtues to explain that something makes you sleep is not very satisfactory. In other words, the account would make natural selection explanations vacuous. Of course, inferences from the past to the present (or future) are ubiquitous in science,[5] but that cannot be the whole story. The point is that for an explanation of the changes in phenotype or genotype/gene frequencies to be satisfactory (i.e. non-vacuous), the *explanans* has to appeal to something more than past changes in those magnitudes. Conceptual independence is not enough to solve the problems at hand; *operational* independence is also required. The key question is whether there is some way of *measuring* fitness values that is independent of *current or past* reproductive success, in the same way that one could directly analyze the weight distribution of

[5] The same point applies to inferences from similar cases, which need not necessarily be past states of the same population.

a coin to figure the strength of its propensity to land on heads, independently of past or current series of throws.

One could think that a possible reply to this objection would be to bite the bullet and acknowledge that the PNS is an *a priori* claim. Sober (2011) holds this position, and explicitly compares the PNS with the claim that substances with a dormitive virtue make you sleep (pp. 572-573; see Diez and Lorenzano 2013 for a possible reply to Sober).[6] However, he also holds that applying the PNS particular cases is an empirical task (p. 579)—because both its antecedent and consequent are empirical claims. The problem here is, as was just said, operational circularity. This problem does not arise for natural selection explanations in Sober's account because he thinks that natural selection theory contains source laws (e.g. "if lions were to hunt and kill slow zebras more successfully than they hunt and kill fast ones, then running fast would promote survival", p. 578). Now these laws are also, according to him, *a priori*. However, the operationalization of the antecedent of this law (a clearly empirical task) can be carried out independently of past measurements of reproductive success and past survival, unlike in the dormitive virtue case. There would be an operational circularity problem with the latter case if the only way to operationalize the antecedent in "substances that have a dormitive virtue make you sleep" was seeing if the substance in question makes one sleep. This operational problem would be independent of the aprioricity status of the general law, considered abstractly. The PIF, however, does not propose that there are source laws such as the one just cited. Thus, they

[6] We thank an anonymous reviewer for bringing this possible response to our attention.

must give a different account of why "fitter organisms tend to have a greater reproductive success" does not incur in the kind of operational circularity just described (i.e. how fitness can be operationalized independently of reproductive success).

In their reply to Rosenberg, Brandon and Beatty acknowledge all of this, and consider that there is such a way:

> We believe that an adequate definition [of fitness] should provide a general specification that is conceptually independent of actual reproductive success. Otherwise, natural selectionist explanations of differential reproduction would be impossible. *An adequate definition should also be operationally independent, i.e., there should be some means of determining fitness differences independent of differences in actual reproductive success*. (…) We certainly need means of determining fitness differences that are independent of actual reproductive success, if differences in actual reproductive success are to be attributed, in turn, to fitness differences. An adequate explanatory account requires evidence in its favor besides the phenomena that it serves to explain. For instance, in the case at hand, the claim that there are fitness differences between the members of a population requires evidence in its favor besides the differences in actual reproductive success that it supposedly serves to explain. *Fortunately, there is such additional evidence, in the form of so-called ecological "optimality" analyses* (Brandon and Beatty 1984, pp. 344-45, emphasis added)

Although discussions on fitness have somewhat moved away from these issues (as the propensity account has become more widely accepted), this idea is still present in the literature. For instance, in a recent article, Millstein claims that:

> With the propensity interpretation, if we seek to explain why type A is leaving a higher number of offspring than type B, the explanation 'because A is fitter than B' means that A has a greater propensity than B to survive and reproduce in the given environment, which means that the physical properties of A in its environment are what cause it to tend to have greater reproductive success than B (with its physical properties). *The relative physical abilities can be determined by engineering optimality models* or other examinations of the physical properties of the organism in its environment, and then be confirmed by measurements of actual descendant contributions. (Millstein 2016, p. 606, emphasis added)

However, despite the idea remaining present, no one has yet presented a detailed account of how these independent measurements of fitness (through optimality models) are supposed to work. To be complete, the PIF still requires an account of how it is that fitness can be assigned independently of actual reproductive success. Our goal in this article is to examine one of the few (if not the only) suggestion that proponents of the PIF did provide: that fitness can be

measured independently of actual reproductive success via optimality models. In the next section, we introduce an example of an optimality model, and use it to outline a possible way in which these measurements could be thought to take place, which also coincides with the way that optimality models are naively portrayed in the optimality literature.

3. Optimality Models and the Measurement of Fitness Values

As we saw above, proponents of the PIF have suggested in various places that fitness values can be measured through optimality models. But how exactly are these ways of measuring supposed to work? The authors do not usually go into much detail. The only developed account we found (within the PIF literature) is that of Beatty (1980). Therefore, in what follows, we consider the example Beatty introduced in that paper, which is taken from Holling (1964).

What Holling sought to determine was the size of the prey to which praying mantids would be most responsive. He reasoned as follows. In general, whenever the amount of energy needed to hunt a smaller and a larger prey is the same, a predator should prefer the larger prey to the smaller one since it would return more energy per unit of energy invested in the mantid's hunt. However, a prey too large may be less likely, or even impossible, to catch (not to mention that tackling a prey too large may put the predator itself in danger). In consequence, there will usually be some trade-off by which the predator will tend to hunt the largest prey it can securely

get hold of. In the case of mantids, Holling took as the *currency* or *optimality criterion* (the unit that the predator should maximize) the amount of energy that a prey of a given size contains (operationalized via the size of the prey). The *constraint* that he considered (the factor that limited the size of prey that the mantids should attempt to hunt) was derived by looking at their grasping mechanism.

The grasping mechanism of the mantid is basically a pincer consisting of two straight arms (the tibia and the femur). Since the pincer tends to force objects outside of the mechanism (see figure 1), the mechanism also includes a hook to secure them in place, as well as a number of spikes that generate friction and counter the various forces present (see Holling 1964, for a more detailed description of this).

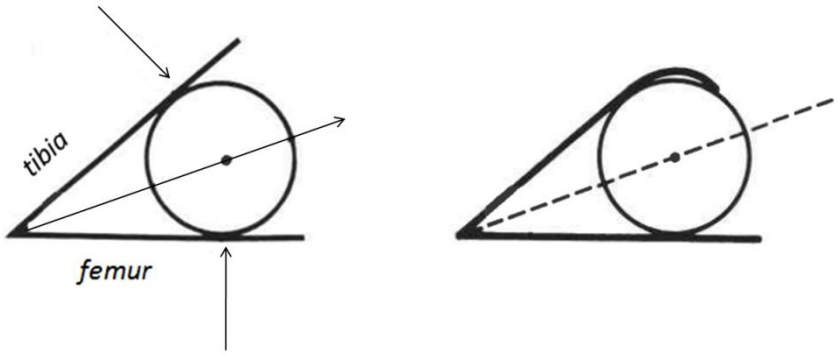

Figure 1. Composition of forces pushing the object out of the mechanism (left); hook holding it in place (top right). Taken from Holling (1964) and modified.

Given this representation of the grasping mechanism's design, Holling reasoned that the largest prey the mantids could safely capture would be the ones whose center passed through the line

that connects the tip of the hook with the femur, when held (see figure 2; as an idealization, the prey are considered to be approximately circular). Any prey larger than that would escape.

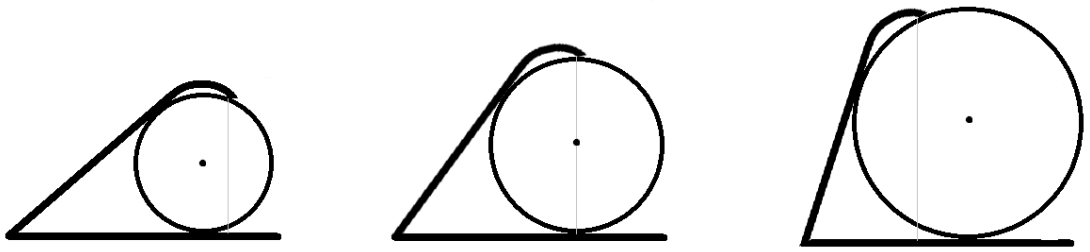

Figure 2. Prey of three different sizes in the grasping mechanism. The one on the left (its center is to the left of the line that connects the hook with the femur) will not escape; the one in the right (its center is to the right of that line) will escape; the middle one (center exactly on the line) is the limiting case.

With this in mind, a simple trigonometric construction can yield the optimal diameter of the prey (the length of the grey line in the middle case of figure 2, relative to the length of some measurable values, such as the length of the tibia). The construction is illustrated in figure 3 below:

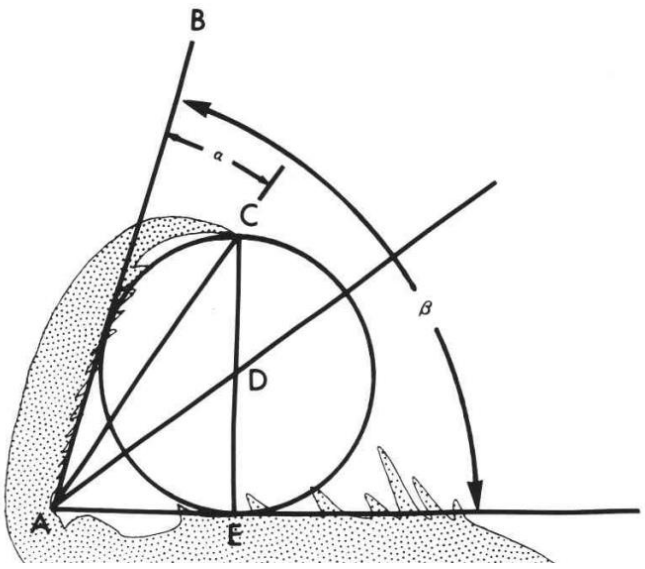

Figure 3. A way to calculate the optimal diameter of the prey. Taken from Holling (1964).

A simple derivation shows that the length of this diameter (line CE in figure 3) is equal to AC × *sin*(β−α). Holling then tested real mantids by showing them prey of different sizes, and they did attempt to hunt prey quite close to the size the model considered optimal.

More generally, an optimality model contains three components or "parts" (Maynard Smith 1978; Parker and Maynard Smith 1990). These are (*a*) a trait-type that can assume a certain range of values (for example, size of prey); (*b*) a "currency" or variable being optimized (such as energetic return from catching such a prey); and (*c*) a set of constraints, possibly of different kinds (physical/mechanical, environmental, developmental, etc.; in the case above, they are physical/mechanical).

The connection between these models and fitness is usually thought (even independently of the PIF account, see below for some examples) to exist through the models' *currency*. One way to understand this is to consider an objection the optimality approach repeatedly received in its inception: that with sufficient imagination to come up with a suitable optimality criterion, any trait whatsoever can be considered optimal (see, for example, Gray 1987). The usual response claims that there are restrictions on the choice of an acceptable currency. The most frequent choice of restriction put forth is that an acceptable currency must either be the fitness of an organism, or correlate with/be a proxy for/be a substitute of/etc. fitness (the precise term used varies from author to author). For example, when characterizing optimality models, Potochnik claims that:

> Optimality models proceed according to the following schema. One determines the range of possible trait values for the phenotype of interest *and the fitness function* that relates these phenotypic trait values to their success in the present environment. Based on this information, an optimality model predicts which trait value(s) would predominate in the population as a result of selection, given enough time in the current environment for an equilibrium to be reached. (Potochnik 2009, pp. 184-185, emphasis added)

The same idea can be found in the biology literature:

> [T]he [optimality] approach entails starting with a specification of those behavioural strategies that can be adopted, *together with an appropriate measure of fitness*. The dependence of fitness on the behavioural strategy adopted is then quantified. Finally, a suitable mathematical optimisation technique such as dynamic programming is used to find the strategy *that maximises fitness.* (McNamara et al. 2001, p. 414, emphasis added)

Of course, a parameter such as energetic return (the currency in Holling's model above) is not usually thought of as a component of fitness (at least not directly). Hence, what is usually claimed is that the chosen currency must be *a proxy* or *substitute* of fitness. For example:

> Biological optimality models aim to represent the available phenotypic strategies along with the constraints and tradeoffs involved in the selection of a trait. Once these components of the model are specified, the optimality modeler can deduce which of the available strategies will optimize the criterion of the model. *Ideally, in biological contexts, this criterion will be fitness (or inclusive fitness), but in most cases a more easily measured proxy is used*—e.g. average energy intake. (Rice 2012, p. 687, emphasis added)

In his article, Beatty took this position as well:[7]

> Ideally, the design problem expresses fitness as a function of particular design variables. However, variables which vary proportionately with fitness may be *substituted* for fitness. For example, the design problem of Holling's model expresses energy efficiency as a function of prey-size choice. (Beatty 1980, p. 535, emphasis added)

If all of this were correct, a possible way of measuring relative fitness values from an optimality model would be to look at the values that its "fitness function" assigns to the different traits. Alternatively, if fitness is not used directly as a currency, relative fitness values could be inferred from the fitness substitute distribution. The following is an example of how that inference might work. Suppose that Holling's model calculates that the energetic returns for hunting a prey of a given size are as in table 1:

[7] Note that, even though Beatty takes this position, proponents of the PIF do not need to claim that, *in general*, optimality models contain a proxy for fitness as currency. All they need to claim is that fitness values are measurable through optimality models *in at least some cases* (not necessarily all of them).

Prey size	Energetic return
1cm	10 cal
2cm	20 cal
3cm	30 cal
4cm	40 cal
5cm	0 cal

Table 1. Example of a possible currency distribution for five trait values

Then, in a population of mantids (for simplicity, let's say one in which individuals have identical size, and in which each individual hunts for only one size of prey), the relative fitness distribution for those four behavioral traits will be w(1cm) = 0.25, w(2cm) = 0.50, w(3cm) = 0.75, w(4cm) = 1 and w(5cm) = 0. These values were obtained simply by dividing each currency value by the greatest currency value (40 cal), so that the currency distribution is transformed into a *relative* currency distribution. This relative currency distribution is then identified with the relative fitness distribution (the idea being that one returns from the values of the "substitute"

quantity to the values of the substituted one).[8] However, the next section shows that this way of considering things is untenable.

4. Why Relative Currency Distributions Cannot be Identified with Relative Fitness Distributions

As shown above, to avoid the problem of explanatory circularity, the PIF requires that the concept of fitness be operationally independent from actual reproductive success. The suggestion made by proponents of the PIF is that fitnesses can be independently measured through the use of optimality models. However, a detailed account of how these ways of measuring are supposed work has not yet been given. In section 3 we provided what we believe to be the most intuitive or obvious account, which is also in accordance with how the connection between fitness and currency is portrayed in the optimality literature. And in this section, we

[8] We should also note that we are assuming (for the sake of simplicity and consistency) that we are in a situation like that described by Holling, where the only relevant prey variable is size. There are other, more complex models in the optimal foraging theory literature (for instance, Schoener 1974) that incorporate more variables. For instance, in Schoener's model, larger prey may be less abundant and more difficult to catch, and thus a predator could prefer smaller prey (even if it is mechanically possible for it to catch the larger prey).

will criticize this account. To do so, we offer two independent lines of reasoning. The first one shows that for the procedure outlined in section 3 to work, fitness and currency distributions must correlate *linearly*. However, since this is almost never the case, the procedure fails. Our second line of reasoning returns to the issue of circularity. We show that the appeal to optimality models as independent ways of establishing fitness differences does not, after all, avoid circularity.

To illustrate the first of these two points, we may consider the following example. Let $c(x)$ be the currency function from an optimality model (i.e. a function that relates a certain trait value x to a currency value), the currency being different from fitness, and let $w(x)$ be the "real" fitness function for the organisms that possess such traits. To make our example more concrete, let the relevant traits x be the different diets of a mantid (where each diet consists of prey of a single size), let $c(x)$ be the average energetic return of such diet, according to Holling's model, and let $w(x)$ be the fitnesses of the mantids that eat said diets. Let it also be the case that the distributions of c and w are as shown in figure 4 (the squares representing fitness values, the crosses currency values):

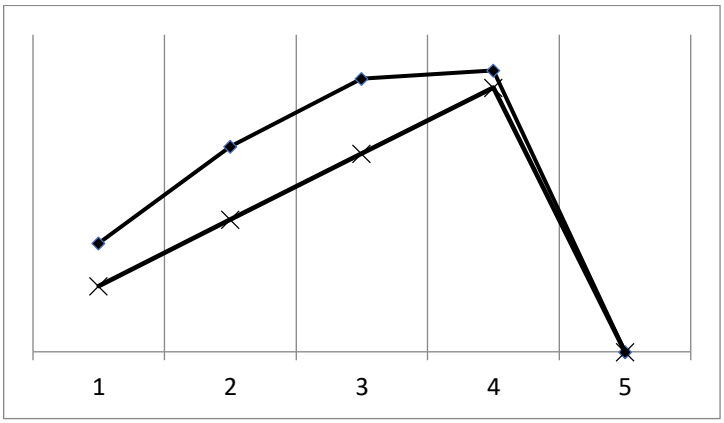

Figure 4. Hypothetical distributions of fitness (squares) and currency (crosses) values for five diets consisting solely of prey of a given size (represented in the horizontal axis). Up to prey of size 4, both distributions increase; however, there are diminishing returns on increasing energy consumption. Prey of size 5 are simply too large to eat, so an individual which exclusively consumes them has both currency and fitness of 0.

Now, it is clear that, in this case, the optimal phenotype (according to the currency function *c*) and the fittest one (according to function *w*) coincide. As Parker and Maynard Smith have noted, every time both functions (currency and fitness) correlate, this will be the case (Parker and Maynard Smith 1990, pp. 28-29). That is, if every time one augments (or diminishes) the other

one augments (or diminishes), and vice versa, then the fittest phenotype and the optimal phenotype will coincide, regardless of the exact shape that both functions possess.[9]

However, this alone does not allow us to use currency values as quantitative measurements of fitness in natural selection explanations. It is not enough that the highest points in both distributions coincide in the same trait value; there must be a sense in which the whole distribution coincides. The reason for this is obvious: in the above example, *currency differences* between organisms that possess different traits do not reflect *fitness differences* between them.[10] If, for example, we were to introduce those currency values into population genetics selection coefficients, we would obtain incorrect predictions for the next generation.

Let us look at a numerical example of this. Consider now a population and a trait-type for which there are two different phenotypes, which we will call D_1 and D_2 (we can think of them, for example, as two diets, containing two different prey sizes). To make the case as simple

[9] In fact, Parker and Maynard Smith show that, in some special cases, when the correlation between functions is not linear (see what follows below) the optimum value and the fittest value may not coincide. For simplicity, we will leave this problem aside. We will show that even if one makes this assumption, there are still problems in thinking that currency values can be directly used to measure fitness values.

[10] One might reply here that, although quantitative fitness differences cannot be established, a qualitative ordering or ranking can be. This point will be addressed in the next section.

as possible, let us also assume that dietary behaviors are completely determined genetically, by a single *locus*, and that these organisms are haploid and reproduce by cloning. Gene A_1 determines diet D_1, while gene A_2 determines diet D_2. Now suppose that A_1's fitness (understood as probability of survival to adulthood[11]) is 0.75, while the energy obtained from prey D_1 is 20 calories per unit consumed. For A_2, those values are 0.5 and 10. So, in this example, as before, the optimal trait and the fittest trait coincide, and currency and fitness distributions (consisting in only two values in this simple example) correlate. Observe now what happens if differences in currency values are used as relative fitness values. The "real" relative fitness values are 1 and 0.66, respectively, for A_1 and A_2, while their "relative fitnesses" (obtained via currency values by the procedure mentioned at the end of section 3) are 1 and 0.5. If we assume that, at generation n, A_1 has a frequency of $p = 0.5$, the "real" prediction is that in the next generation $p' = 0.59$ (*ceteris paribus*, assuming no mutation, migration, drift, etc.). However, the prediction derived using currency values as fitness values is $p' = 0.66$.

[11] We are utilizing a very basic textbook population genetic model, in which fitness is interpreted as the probability of survival to adulthood. For our criticism to be more accurate, we should associate each diet with a probability distribution of leaving n offspring, and then some mathematical way of converting each distribution into a scalar. The first option was chosen for the sake of simplicity and clarity.

In order for both predictions to coincide, and in order for us to be able to claim that an optimality model contains a "fitness function" or a variable that can be *quantitavely used as its substitute*, we would need both variables (currency and fitness) to correlate *linearly*; that is, equal increases or decreases in one function (*c* or *w*) would have to correspond to equal increases or decreases in the other. So, whenever one wishes to measure fitness values from an optimality model (in this manner, by inferring them from relative currency differences), it must be the case that the currency of that model and fitness correlate linearly.

However, there are reasons to think that in many models (for instance, those pertaining to optimal foraging theory, or OFT for short), this will tend not to be the case. The reason we should not *a priori* expect linearity to be the case can be seen from looking at biological practice and the way these models are tested. The process typically goes as follows. First, a trait is identified, and a problem is formulated (for instance, "what is the optimal diet of a predator in *X* and *Y* circumstances?"). Then, an optimization criterion is devised (in OFT, usually one that is not a component of fitness), and a number of constraints are identified. With all of this in mind, a mathematical model is formulated, and an optimum is derived. After this is done, real populations of organisms are checked to see if they possess the theoretically optimal trait or not. If not, then it may be a sign that some assumption in the model is not correct, or simply evidence that the real traits are not optimal. If, however, the optimal value is found, then the model is thought to be successful, regardless of whether or not the entire distribution of currency values correlates linearly with the distribution of fitness values.

What this shows is that the success of a model has, in practice, much weaker conditions than linear correlation. Within the literature on OFT, in the few cases where correlation with fitness was actually tested, a correlation was found but it tended not to be linear. In most cases, however, linearity has not even been tested (see Schoener 1987, p. 56; Sih 1982).

Furthermore, in many cases, it could even be impossible to test for correlation with fitness. For example (now outside of OFT), Labouygues and Figureau (1984) studied the reason why, for aminoacids codified by more than one codon, variation in the code tends to occur in the same position (typically the third). Their explanation is based on optimality. The reason for the typical placement of the variation is that this minimizes the effects of punctual mutations (the change in the aminoacid coded). They compared the actual codes for each aminoacid with other possible codes. As a result, they found that, for example, the codes for the five aminoacids that are codified by four codons are included in a group of 48 optimal codes, out of more than 630.000 possible ones. To do this, they did not perform any kind of evolutionary study about the differential fitness of extant organisms that have different codes. There might be reasons to think that an organism with a code that is less resistant to mutations would be less fit, but these reasons were not, and very likely cannot, be tested. Considering all this, the systematic occurrence of exact linear correlation between fitness and currency would be too much of a coincidence.

As a side note, all of this also shows that the more general position outlined in section 2, according to which *every* successful optimality model must contain a proxy for fitness as

currency, is also not adequate (though we noted that the PIF does not need to claim this, see footnote 3). That is, a model could in principle be considered successful (in the sense of the paragraph from above) without its currency linearly correlating with fitness. And since linear correlation with fitness is a necessary condition for a currency to act as a fitness proxy, then it could also be the case that a model is successful without having a fitness proxy as its currency.[12]

In sum, the propensity account requires that we be able to measure relative fitness differences independently of actual reproductive success. However, in the cases at hand, the differences in currency values very likely do not coincide with the differences in fitness values. Thus, the currency in Holling's model (and in most OFT models) is not fitness, nor a proxy for it, and fitnesses cannot be quantitatively established from it, in the way the propensity account requires.

Our second line of argumentation shows that it is not that clear how the appeal to optimality models allows us to avoid circularity. We show this via a disjunction elimination. That is, either every optimality model contains fitness (or a proxy for fitness) as its currency, or this is not the case (we have already shown the first disjunct to be inadequate, but for the sake of argument—to make our two lines of reasoning truly independent—let us assume we did not).

[12] For a similar conclusion regarding OFT, although argued from different premises, the reader may see Bolduc & Cézilly (2012, pp. 860-861).

We will show that whether or not every optimality model contains fitness as its currency, circularity seems to reappear.

Firstly, if not every optimality model contains a proxy for fitness as its currency, then not every optimality model is relevant for measuring fitness values (in the way suggested in section 3). Thus, *before* inferring fitness values from an optimality model, we must know that the model we are looking at is a relevant one. But knowing that requires knowing that its currency is in fact a proxy for fitness. And knowing that requires prior knowledge of fitness values (the very thing we are trying to establish). So, if not every model has a proxy for fitness as its currency, then independent measurements of fitness through optimality models are circular because the choice of a relevant model from which to infer fitness values requires prior knowledge of fitness values.[13]

On the other hand, if every model must have a proxy for fitness as its currency, then the particular model we choose is not so relevant. However, circularity appears somewhere else. If the way to appropriately choose a currency (an optimality criterion) is to either choose fitness

[13] That the currency is a good proxy for fitness could perhaps be inferred from similar cases, not necessarily from the same population, or even the same species. However, as stated in footnote 5, we would still need to have measured the fitnesses in that other population or species beforehand. Thus, we would not have a measurement of fitness that is independent of actual reproductive success.

itself or a proxy for fitness, then the choice of an appropriate currency is what presupposes prior knowledge of fitness values. That is, the construction of the model itself requires knowledge of fitness in order to choose a currency. In other words, independently estimating fitness requires building an optimality model, but building the model itself requires prior knowledge of fitness.

In summary, whichever option one chooses (every optimality model contains a fitness proxy as currency, or not), establishing fitness values by directly inferring them from relative currency values requires prior knowledge of fitness values, either to choose an appropriate model or to choose a legitimate currency in the construction of the model itself.

Having said all this, we do not believe that the idea that optimality models are somehow relevant for fitness estimations is wrong. The disjunction elimination argument presented above is too strong. Scientific theories do tend to have some (non-vicious) circularity because they sometimes propose new terms to explain phenomena, while at the same time such phenomena provide the criteria for the application of those terms, through the laws or principles that contain them. In the following section, we will argue that some such principles linking optimality and fitness are (implicit) in the practice of evolutionary biology, and thus that the mere appeal to propensities is not enough. This link does exist in the practice in evolutionary biology, and the question is how best to reconstruct it, in general, and how to develop the suggestion made by the PIF proponents, in particular. In the next section, we outline a metatheoretically more adequate way to capture this fact.

5. Optimality and Fitness Reconsidered

Thus far, our argumentation has been mostly destructive. In this section, we introduce a general picture of what we believe to be a more adequate account of the relation between optimality models and fitness.

The key to this is to return to what we had presented as a paradigmatic example of a successful application of a propensity-based account: the landing of a coin. As we said, a way of measuring the probability (propensity) of a coin landing heads, that is independent of past throws, would consist in looking directly at its physical properties (e.g. its weight distribution). But this would not be enough since, of course, an analysis of a weight distribution does not, by itself, imply any landing probabilities. We would also need to pay attention to the physical theories or principles that connect weight distributions with landing probabilities. In the same way, an optimality analysis does not, by itself, imply any fitness differences between organisms. Our arguments from the last section show that currency and fitness are *conceptually* distinct, since they can assume different values in a given case. Therefore, we need something else, some principle(s) that allow us to go from one to the other.

What the standard optimality literature implicitly assumes as a principle is something along the lines of "differences in currency quantitatively correspond to differences in fitness"

(see the usual characterization of optimality models above). Section 4 has shown that this principle is too strong, and that we have no reason to believe it to be true.

Alternatively, one could think of a weaker principle, that posits only a (possibly non-linear) correlation between optimality and fitness. In that way, determining currency values in an optimality model could allow us to establish a fitness *ranking* between different kinds of organisms, though not their precise fitness *values*. Some authors (Casanueva 2011; Ginnobili 2016) have tried to make this connection more explicit by introducing a second conditional to the PNS, which goes from the optimality / environmental effectiveness of a trait to fitness. For example, Ginnobili formulates the PNS as follows:

> Organisms that possess a trait that performs more effectively in a given environmental context tend to be fitter than organisms that do not possess it, tending to have (if the trait is heritable) a greater differential reproductive success than those other organisms. (Ginnobili 2016, p. 18, slightly modified to fit our terminological uses)

The point would be, again, that the connection between trait effectiveness and fitness is factual/empirical. Optimality models (although Ginnobili does not specifically refer to them) directly measure something else, not fitness, and there is no requisite for them to have a fitness proxy as currency. Qualitative or comparative determinations of fitness through optimality

models occur via this principle, they do not stem from an *a priori* requirement of currency and fitness correlation for building the model (hence avoiding the circularity problem from above).

An interesting consequence of this reconstruction of the theory of natural selection, if adequate, is that PIF would not provide a complete reconstruction of PNS but only of one of its parts. We will return to this point in the next section.

However, some might argue that, in order to fully avoid circularity charges, the PIF needs more than qualitative / comparative independent determinations of fitness. That is, since it identifies fitness with a *probability*, and since probabilities are quantitative variables, a *fully* independent operationalization of fitness would have to be quantitative in nature. In other words, even though proponents of the PIF mostly discussed the circularity issue in relation to the PNS, this account purports to do more than show that the PNS in non-circular. In particular, it also purports to account for the uses of the concept of fitness within population genetics. As Pence and Ramsey recently put it:

> 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. (Pence and Ramsey 2013, p. 852)

This is important because, again, within population genetics, fitness is a quantitative concept (it is represented as a set of probabilistic coefficients). This theory is also quantitative in the sense that it predicts/explains *numerical* variations in genotype/gene frequencies, and does so (at least partially) in function of the *values* of the fitness coefficients. Therefore, showing that the PNS, as it is specified (in Brandon's terminology "instantiated") in particular population genetic selection models, is operationally non-circular/vacuous, requires that fitness differences can be *quantitatively* measured independently of reproductive success. Optimality models were good candidates for establishing this connection in the first place precisely because they are quantitative in nature. The is still, then, an open puzzle for the PIF, and one for which their proponents must still provide an account.

In the following section, we draw some more general implications from this analysis, regarding the status of probability theory and optimality analyses in evolutionary biology.

6. The nature of the PNS and Evolutionary Biology

If the PIF advocates were correct, although not tautological or trivial, the PNS would have a peculiar status, since all it would do is relate expected frequencies to actual frequencies. Brandon (one of the original proponents of PIF), has focused on this point the most. According to him, the PNS is an instantiation of what he calls the "Principle of Direct Inference" from Probability Theory (Brandon 1978; 2006, p. 333), which is the principle that allows us to infer (or predict) actual frequencies from probabilities. This would imply that Natural Selection Theory has a peculiar status as a scientific theory, which is in consonance to his wider views about probability theory being the reductive foundation for all evolutionary biology (McShea and Brandon 2010, pp. 108-109).

If what we held in the previous section is adequate, then certain limits to this idea can be pointed out. A strong criticism, assuming the tripartite presentation of the PNS, would be to hold that the PIF does not succeed in reconstructing natural selection theory in its entirety, since it only focuses on one of the parts of the PNS, leaving aside the part that links the differences in optimality with differences in fitness. Adopting a more complex version of PNS would show that it is more than just the principle of direct inference, and that evolutionary biology is more than just applied probability theory.[14]

[14] Notice that we are not claiming here that the PIF advocates are mistaken *because* the PNS is not *a priori* or analytical. What we did show in the previous sections is that their own claim that fitness has to be determinable independently from fitness, through the use of optimality models,

A weaker criticism can be made, which does not question PIF account in general, but rather only the additional thesis of the reducibility of evolutionary biology to probability. This would consist in pointing out that the determination of fitness through optimality presupposes factual principles that are essentially evolutionary (whether they are weak and comparative or of a more quantitative nature). In this sense, it would be shown that, although PIF is an adequate explication of the concept of fitness, it does not imply that evolutionary biology is nothing but applied probability. Once again, one could appeal to Sober and claim that these evolutionary principles (the source laws in his account) are *a priori*—at least when considered abstractly—but even then they would be more than purely applied probability (for an example, see the source law about lions and zebras cited above).

7. Conclusions

In this article we have developed in more detail the proposed connection between the PIF and optimality models. We have shown that in order to avoid the issues of explanatory circularity or vacuousness, this account requires not only that fitness be conceptually independent from

---

should lead them to reject the idea that evolutionary theory is nothing more than applied probability (because, as we showed, if this inference can be made, then it is much more indirect than thought and presupposes more than pure probability and the direct inference principle).

reproductive success, but also operationally independent from it. That is, the PIF requires that fitness values can be measured independently of actual reproductive success (either past success in the same population, or in similar populations). The way proponents had suggested this can be done is through optimality models. However, these proponents never set out to establish, in detail, how this independent determinations or measurements are supposed to work.

The destructive part of our article aimed to show the way in which most of the optimality literature portrays this connection is inadequate, since quantitative relative fitness values cannot be directly extrapolated from distributions of currency values, at least in an important subset of cases. The reason is that this would yield incorrect fitness values since (in practice) currencies in successful optimality models are not required to correlate linearly with fitness. We have also argued, independently, that this proposal would also lead us back to circularity, because either the choice or the building of an appropriate model from which to infer fitness values would presuppose prior knowledge of fitness values.

On the other hand, our positive proposal begins by noting that optimality (or currency) and fitness are two distinct concepts (there is no definitional or "analytic" equivalence between them), and thus any proposed connection between them must be made through some empirical principles. We argued that a weaker principle to the one implicitly assumed in the optimality literature can allow us to get fitness rankings from currency values.

Both these points allow us to point out certain limits to the PIF. The sole appeal to propensities is not enough to explicate the role that optimal studies have in evolutionary biology

or to satisfactorily avoid the problem of tautologicity. For that, it is necessary to either make explicit the empirical principles that link optimality with fitness, allowing independent determinations of fitness, or to give more complete versions of PNS. As a corollary, we have shown that the idea that evolutionary biology is nothing more than probability theory is incorrect.

References

Beatty J (1980) Optimal-Design Models and the Strategy of Model Building in Evolutionary Biology. Philos Sci 47(4):532-561

Bolduc J-S, Cézilly F (2012) Optimality Modelling in the Real World. Biol Philos 27(6):851-69

Brandon R, Beatty J (1984) The Propensity Interpretation of 'Fitness'--No Interpretation Is No Substitute. Philos Sc 51(2):342-47

Brandon R (1978) Adaptation and Evolutionary Theory. Stud Hist Phil Sc 9:181-206

Brandon R (2006) The Principle of Drift: Biology's First Law. J Philos 103(7):319-335

Casanueva M (2011) A Structuralist Reconstruction of the Mechanism of Natural Selection in Set Theory and Graph Formats. In Martinez J, Ponce de León A (ed) Darwin´s Evolving Legacy. Siglo XXI, México, pp 177-192

Darwin CR (1859) On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. John Murray, London

Díez J, Lorenzano P (2013) Who Got What Wrong? Fodor and Piattelli on Darwin: Guiding Principles and Explanatory Models in Natural Selection. Erkenntnis 78:1143–1175

Fisher R (1930) The Genetical Theory of Natural Selection. The Clarendon Press, Oxford

Fodor J, Piattelli-Palmarini M (2010) What Darwin Got Wrong. Farrar, Straus and Giroux, New York

Futuyma DJ (2005) Evolution. Sinauer Associates Inc, Sunderland, Massachusetts

Ginnobili S (2016) Missing Concepts in Natural Selection Theory Reconstructions. Hist Philos Life Sci. https://doi.org/10.1007/s40656-016-0109-y

Gray RD (1987) Faith and Foraging: A Critique of the 'Paradigm Argument from Design.'. In Kamil AC, Krebs JR, Pulliam HR (eds) Foraging Behavior. Springer US, pp 69-140

Holling, CS (1964) The Analysis of Complex Population Processes. The Can Ent 96(1-2):335-347

Labouygues J-M, Figureau A (1984) The Logic of the Genetic Code: Synonyms and Optimality against Effects of Mutations. Orig Life 14(1):685-92

Luque VJ (2017) One equation to rule them all: a philosophical analysis of the Price equation. Biol Philos 32(1): 97-125

Maynard Smith J (1978) Optimization Theory in Evolution. Annu Rev Ecol Syst 9:31–56

McNamara JM, Houston AI, Collins S (2001) Optimality Models in Behavioral Biology. SIAM Rev 43(3):413-466

McShea D, Brandon R (2010) Biology's First Law. The University of Chicago Press, Chicago

Mills SK, Beatty JH (1979) The Propensity Interpretation of Fitness. Philos Sc 46(2):263-286

Millstein RL (2016) Probability in Biology: The Case of Fitness. In Hájek A, Hitchcock CR (eds) The Oxford Handbook of Probability and Philosophy. Oxford University Press Oxford, pp 601-622

Mivart GJ (1898) The Groundwork of Science. John Murray, New York

Parker GA, Maynard Smith J (1990) Optimality Theory in Evolutionary Biology. Nat 348:27-33

Pence CH, Ramsey G (2013) A New Foundation for the Propensity Interpretation of Fitness. Br J Philos Sci 64(4):851-81

Peters RH (1976) Tautology in Evolution and Ecology. Am Nat 110(971):1-12

Popper KR (1974) Unended Quest: An Intellectual Autobiography. Open Court, La Salle

Potochnik A (2009) Optimality Modeling in a Suboptimal World. Biol Philos 24(2):183-97

Rice C (2012) Optimality Explanations: A Plea for an Alternative Approach. Biol Philos 27(5):685-703

Ridley M (2004) Evolution - Third Edition. Blackwell, Malden.

Rosenberg A (1982) On the Propensity Definition of Fitness. Philos Sci 49(2):268-273

Rosenberg A, Bouchard F (2015) Fitness. In: Zalta EN (ed) The Stanford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/fitness/. Accessed 13 August 2019

Schoener TW (1974) The Compression Hypothesis and Temporal Resource Partitioning. Proc Natl Acad Sci U S A 71(10):4169-4172

Schoener TW (1987) A Brief History of Optimal Foraging Ecology. In Kamil AC, Krebs JR, Pulliam HR (eds) A Brief History of Optimal Foraging Ecology. Springer US, New York, pp 5-67

Sih A (1982) Optimal Patch Use: Variation in Selective Pressure for Efficient Foraging. The Am Nat 120(5):666-685

Sober E (2011) A Priori Causal Models of Natural Selection. Australas J Philos 89:571–589

Sober E (2013) Trait fitness is not a propensity, but fitness variation is. Stud Hist Philos Sci Part C Stud Hist Philos Biol Biomed Sci 44:336–341

Vallejo F (1998) La Tautología Darwinista y Otros Ensayos de Biología. Taurus, Madrid