Empirical adaptationism revisited:

Is it testable and is it worth testing?

Mingjun Zhang （张明君)

School of Philosophy, Fudan University, Shanghai, China

Email: mingjunzhang@fudan.edu.cn

Abstract:

Empirical adaptationism is often said to be an *empirical* claim about nature, which concerns the overall relative causal importance of natural selection in evolution compared with other evolutionary factors. Philosophers and biologists who have tried to clarify the meaning of empirical adaptationism usually share, explicitly or implicitly, two assumptions: (1) Empirical adaptationism is an empirical claim that is scientifically testable; (2) testing empirical adaptationism is scientifically valuable. In this article, I challenge these two assumptions and argue that both are unwarranted given how empirical adaptationism is currently formulated. I identify a series of conceptual and methodological difficulties that makes testing empirical

adaptationism in a biologically non-arbitrary way virtually impossible. Moreover, I show that those in favor of testing empirical adaptationism have yet to demonstrate the distinctive value and necessity of conducing such a test. My analysis of the case of empirical adaptationism also provides reasons for scientists to reconsider the value and necessity of engaging in scientific debates involving the notion of overall relative causal importance.

Keywords: Empirical adaptationism; overall relative causal importance; natural selection; testability; optimality model; representative sample

1. Introduction

The debate about adaptationism has been one of the central topics in evolutionary biology and the philosophy of biology (Gould and Lewontin 1979; Mayr 1983; Godfrey-Smith 2001; Lewens 2009; Orzack and Forber 2010). Instead of having a single, unified meaning, it refers to a family of views concerning the causal, methodological, or explanatory importance of natural selection in the course of evolution or in evolutionary research (Godfrey-Smith 2001; Lewens 2009; Orzack and Forber 2010). One variety of these adaptationist views tries to make an empirical claim about the *overall relative causal importance* of natural selection in evolution compared with non-selective evolutionary factors. In its most general form, it states that natural selection is, in most cases, the most (or the only) important cause of evolution. Since this view is intended to make an empirical claim about nature, I will, for the sake of discussion, follow Godfrey-Smith (2001) and call it "empirical adaptationism."[1]

_____

[1] There are different ways to name this kind of adaptationist view in the literature. For example, Sober (1998) calls it "a non-trivial empirical thesis about the history of life"; Godfrey-Smith (2001) uses the term "empirical adaptationism"; Lewens (2009) further distinguishes between three forms of empirical adaptationism, and regards the view introduced here, which he calls "pan-selectionism," as merely one form of empirical adaptationism.

A number of philosophers and some biologists have tried to clarify the meaning of empirical adaptationism and define it in ways that do not make it trivially true or obviously false (e.g., Sober 1987; 1998; 2000; Orzack and Sober 1994a; Godfrey-Smith 2001; Lewens 2009). Although these scholars may disagree on the details of how exactly empirical adaptationism should be formulated, many of them usually share, explicitly or implicitly, two assumptions about it: (1) Empirical adaptationism, while its truth is currently unknown or controversial, is an empirical claim about nature that is *scientifically testable* in the long run; (2) testing empirical adaptationism is *scientifically valuable*.

In this article, I will reexamine these two assumptions and argue that both are unwarranted given how empirical adaptationism is currently formulated. In Section 2, I survey some influential formulations of empirical adaptationism in the literature and identify as its main theme the overall relative causal importance of natural selection in evolution. I further distinguish two empirical adaptationist views regarding this theme: (I) Natural selection is, in most cases, the *most important* cause of phenotypic evolution; (II) natural selection is, in most cases, the *only important* cause of phenotypic evolution. In Section 3, I challenge the testability of claim (I) by considering some problematic cases where the evolution of a trait is influenced by both natural selection and a non-selective factor but we are not allowed to privilege one evolutionary factor as causally more important than the other. The aim is to show that in many cases we may not be able to compare the relative causal importance of natural selection and other evolutionary factors in an intelligible way. Sections 4 and 5 challenge the testability of claim (II). I introduce Orzack and Sober's formulation of empirical adaptationism as a representative example of claim (II) and identify two methodological difficulties in testing this version of empirical adaptationism: One concerns the difficulty of meaningfully aggregating testing results

of optimality models with varying backward time scales; the other concerns the difficulty of obtaining a representative sample of phenotypic traits due to the poorly defined statistical population. These issues, if cannot be resolved, will undermine the status of empirical adaptationism as an empirical thesis that is testable in a biologically non-arbitrary way. In Section 6, I consider various types of value that might be attached to testing empirical adaptationism and argue that none of them can be used to justify the necessity of conducting such a test. Hence, those in favor of testing empirical adaptationism have yet to demonstrate the distinctive value and necessity of doing so. I conclude by inviting scientists to reconsider the value and necessity of engaging in scientific debates involving the notion of overall relative causal importance.

2. Empirical adaptationism and the notion of overall relative causal importance

I start by examining some influential formulations of empirical adaptationism in the literature. The aim of doing so is to figure out the *main theme* of empirical adaptationism in its various forms:

> Natural selection has been the only important cause of most of the phenotypic traits found in most species. (Sober 1998, p. 72)

> Natural selection is a powerful and ubiquitous force, and there are few constraints, except general and obvious ones, on the biological variation that fuels it. To a large degree, it is possible to predict and explain the outcome of evolutionary processes by attending only to the role played by selection. No other evolutionary factor has this degree of causal importance. (Godfrey-Smith 2001, p. 336)

> [N]atural selection is the most significant of the evolutionary forces that act on populations. (Lewens 2009, p. 162)

One commonality among these formulations is that they all concern the *overall relative causal importance* of natural selection in evolution compared with other evolutionary factors. The notion of overall relative causal importance is the combination of two simpler notions – *relative*

causal importance and *relative frequency*. The notion of relative causal importance is involved in cases where the production of a phenomenon is influenced by more than one causal factor and different factors may make different amounts of causal contributions to the focal phenomenon. A typical example is the case where a particle is accelerated by two forces acting in the same direction. One of the two forces can be regarded as a more important cause if it makes more causal contributions to the particle's acceleration. The notion of relative frequency is involved in cases where there exist multiple, alternative theories of a general phenomenon but each theory can only explain a proportion of instances of that phenomenon (Beatty 1997; Kovaka 2021). Proponents of different theories may debate which theory covers a larger proportion of instances of the general phenomenon. For example, speciation can occur via different mechanisms, and evolutionary biologists and systematists have argued about the relative frequency of different modes of speciation. The notion of overall relative causal importance is a combination of the above two notions, because it concerns not only whether a factor plays a more important causal role in the production of particular instances of a general phenomenon, but also how often that is the case in the relevant domain.

In the case of empirical adaptationism, in order to evaluate the overall relative causal importance of natural selection in evolution, biologists need to proceed in two steps. First, they need to test whether natural selection is the most or the only important cause of the evolution of various *particular traits*. Second, they need to determine *how often* it is the case that natural selection is the most or the only important cause of the evolution of a trait.

Despite similar wording, the claim that natural selection is the *most important* cause of the evolution of a trait is different from the claim that natural selection is the *only important* cause of the evolution of a trait. The former claim does not deny the possibility of non-selective

5

evolutionary factors being causally important in the evolution of a trait, but simply emphasizes the causal superiority of natural selection over other factors. The latter claim, however, regards the role of non-selective evolutionary factors as negligible in the evolution of a trait. Hence, we can further distinguish two empirical adaptationist views:

(I) Natural selection is, in most cases, the *most important* cause of phenotypic evolution.

(II) Natural selection is, in most cases, the *only important* cause of phenotypic evolution.

Notice that empirical adaptationism can be formulated at both the molecular and phenotypic levels. The testability and correctness of empirical adaptationism at these two levels are not necessarily the same. In this article, I shall limit my discussion to the phenotypic level, leaving the testability and value of empirical adaptationism at the molecular level for another essay. In the following sections, I will introduce claims (I) and (II) in more detail and examine their testability respectively.

3. Conceptual difficulties in comparing the relative causal importance of natural selection and other evolutionary factors

Let us first consider claim (I): Natural selection is in most cases the *most important* cause of phenotypic evolution. The idea that natural selection is in some sense more important than other evolutionary factors has a long history in the development of evolutionary thinking. At the very end of the Introduction in the first edition of *On the Origin of Species*, Darwin claimed that "I am convinced that Natural Selection has been the *main* but not exclusive means of modification" (Darwin 1859, p. 6, my italics). In the fifth edition, he revised the wording and stated that "I am convinced that Natural Selection has been the *most important* but not the exclusive means of modification" (Darwin 1869, p. 6, my italics). Darwin is not alone in this kind of thinking. For

6

example, Jonathan Losos, a leading evolutionary biologist today, thinks that "evolution is the historical occurrence of change, and natural selection is one mechanism – *in most cases the most important* – that can cause it" (Losos 2014, p. 3, my italics).

As Godfrey-Smith (2001, p. 343) rightly points out, to test empirical adaptationism scientifically, "[w]hat is needed is a way of comparing the relative causal importance of natural selection and other evolutionary factors." When a phenomenon is the result of multiple causal factors, a common way of deciding which is the most important is to partition and compare the causal contributions of different factors to the focal phenomenon. In the following, however, I will argue that it is not always possible to do so in the case of empirical adaptationism. My strategy is to consider various types of cases where the evolution of a trait is influenced by both natural selection and a non-selective factor and show that at least in those cases it does not make sense to privilege one evolutionary factor as causally more important than the other.

Case 1: Pleiotropy is the phenomenon that one gene has more than one phenotypic effect. Suppose that two phenotypes always co-occur among the individuals in a given population due to the effect of pleiotropy. One of these two phenotypes has beneficial effects on its carrier᾽s fitness, while the other has deleterious effects (this is called antagonistic pleiotropy). The deleterious phenotype that should have been eliminated by natural selection may be maintained and become fixed in the population because of the stronger positive effects of the beneficial phenotype. In this case, there is no doubt that both pleiotropy and natural selection are important causal factors in the evolution of the deleterious phenotype: The effect of pleiotropy is to link the beneficial and the deleterious phenotypes, while the role of natural selection is to drive both of them to fixation given their overall positive effect. However, it is far from clear how we can compare the relative causal importance of these two evolutionary factors, for we lack a way to

quantitatively apportion the causal contributions made by pleiotropy and natural selection to the evolution of the deleterious phenotype in the given population (Sober 1987, p. 115).

Those who attempt to privilege the causal role of natural selection may respond that natural selection has the power to break the pleiotropic link between two phenotypes. For example, Dawkins (1982, p. 35) has argued that "[i]f a mutation has one beneficial effect and one harmful one, there is no reason why selection should not favour modifier genes that detach the two phenotypic effects." Although every biologist will agree that the breaking of a pleiotropic link is possible, this possibility per se does not guarantee that an existing pleiotropic link will necessarily be broken, nor does it tell us when this link will be broken. Since testing empirical adaptationism involves assessing the *actual* relative causal importance of natural selection and other evolutionary factors, the possibility mentioned by Dawkins does not change the fact that we cannot compare the relative causal importance of pleiotropy and natural selection when a pleiotropic link is still present.

Case 2: In his Shifting Balance Theory, Sewall Wright (1931; 1932) depicts a scenario where a large population is divided into many partially isolated subpopulations and its adaptive evolution is driven by the interaction between natural selection and genetic drift. The whole process can be understood as consisting of three phases (Skipper 2002): In the first phase, genetic drift causes gene frequencies to fluctuate in subpopulations, allowing some of them to move across adaptive valleys and reach the base of a higher adaptive peak; in the second phase, natural selection operates within subpopulations, moving them to local adaptive peaks; in the third phase, organisms from more fit subpopulations migrate to less fit ones, and natural selection between subpopulations increases the average fitness of the whole large population.

It is clear that natural selection and genetic drift play different roles in the above scenario: Genetic drift helps subpopulations shift across adaptive valleys toward higher adaptive peaks; natural selection increases the average fitness of both subpopulations and the whole population. However, it is unclear how we can compare the relative causal importance of natural selection and genetic drift in cases like this. Without genetic drift, subpopulations will be held at the nearest adaptive peaks by natural selection, having little chance to move across the adaptive valleys toward higher adaptive peaks; without natural selection, cumulative adaptive changes are unlikely to occur, no matter in subpopulations or the whole population. Hence, neither of these two evolutionary factors can be said to be causally more important than the other – both play an essential role in the evolutionary process described above.

Case 3: Suppose that initially there exist three variants (V_1, V_2, V_3) of a trait in a population, and their fitness ranks as follows: Fitness (V_1) > Fitness (V_2) > Fitness (V_3). Among these variants, V_1 and V_2 are new variants that have recently emerged in the population, each having a very low frequency. Suppose that V_1 is lost from the population due to the effect of random genetic drift. Among the remaining two variants, V_2 is favored by natural selection and spreads to fixation. In this case, genetic drift eliminates the fittest variant from the population, while natural selection increases the frequency of the second fittest variant. The effect of natural selection is influenced by the effect of genetic drift: If genetic drift had not removed the fittest variant V_1 from the population, natural selection would have driven V_1 instead of V_2 to fixation. Hence, both genetic drift and natural selection are essential to explain why V_2 can eventually get fixed in the population, and there is no intelligible way to compare the relative causal importance between these two evolutionary factors.

The difficulties identified in the above problematic cases are conceptual rather than empirical: In these cases, I have assumed that we are able to clearly measure and distinguish between the effects of natural selection and other evolutionary factors, but this still does not allow us to claim one evolutionary factor as being causally more important than the other.[2] This is because in these cases natural selection and the relevant non-selective factor play *qualitatively different* causal roles in the evolution of the relevant trait, and both are essential for explaining the evolution of that trait. As a consequence, the comparison problem cannot be expected to be solved simply through the advance of technology or more detailed empirical research in the future.

Here is one possible objection to my argument: What the above problematic cases really show is that in cases like these, we cannot compare the relative causal importance of natural selection and other evolutionary factors in a particular way, i.e., by apportioning and comparing their causal contributions to the evolution of a trait in a given population, not that such comparison is impossible in any conceivable sense. I agree. However, if one insists on the possibility of comparing the relative causal importance of natural selection and other evolutionary factors in the above problematic cases, it should be *their* responsibility to make sense of such comparison. To the best of my knowledge, no successful strategies have been devised in the literature on empirical adaptationism to fully address those problematic cases. For example, I have discussed a common response (Dawkins' comment on pleiotropy) to Case 1 and shown that it does not really solve the problem. And it is unclear how the comparison problem in Cases 2 and 3 can be fully addressed.

---

[2] This is not to deny the existence of empirical difficulties in measuring and distinguishing between the effects of different evolutionary factors. In fact, such difficulties are one of the principal sources of controversies in evolutionary research (e.g., see Millstein 2008).

Another possible objection is that the kinds of problematic cases provided above should be *common enough* in order to cast doubt on the testability of empirical adaptationism as formulated by claim (I). I agree that information regarding the commonness of these problematic cases should be provided. But what counts as "common enough"? One way to understand this notion is to think that a type of problematic cases is common when it has multiple instances across a broad diversity of species or higher taxa. Given this understanding, the three types of problematic cases I provided are indeed common. Case 1 concerns the phenomenon of antagonistic pleiotropy. Recent literature reviews (e.g., Austad and Hoffman 2018) suggest that pleiotropic trade-offs between traits enhancing early life fitness and traits shortening lifespan are very common throughout the animal world (and potentially all living domains). Case 2 concerns scenarios where traits evolved as envisaged by Wright's Shifting Balance Theory (SBT). Wade and Goodnight (1991) provide experimental confirmation of this theory using the flour beetle *Tribolium castaneum*. Empirical evidence supporting the SBT is also found in field studies of natural populations of many species in plants, insects (for a review, see Wade and Goodnight 1998) and frogs (Chouteau and Angers 2012). Case 3 concerns the loss of beneficial mutations due to genetic drift. Theoretical population genetic models (Haldane 1927; Kimura and Ohta 1971) have shown that the probability that a rare beneficial mutation is *not* lost by drift is approximately twice its selective advantage. Since most beneficial mutations have low selective advantages, it is reasonable to infer that the majority of newly arisen beneficial mutations are lost owing to genetic drift. These three types of problematic cases are certainly not exhaustive. When taking into account all types of problematic cases, it is easy to find multiple instances across a broad diversity of species or higher taxa.

Those in favor of testing empirical adapatationism may focus instead on the relative frequency, rather than the absolute number, of problematic cases. They may require that problematic cases are common enough only when their overall relative frequency in all cases of phenotypic evolution surpasses a certain threshold such as 50%. This requirement, despite its intuitive appeal, faces a serious methodological challenge. Determining the relative frequency of problematic cases requires a *reasonably representative* sample of all phenotypic traits. However, as I shall argue in Section 5.2, there are serious methodological difficulties in obtaining such a sample (see Section 5.2 for more details), making it difficult to perform a proper estimation of the relative frequency of problematic cases. This fact actually provides another reason to be skeptical of the testability of claim (I): Given the difficulty of performing a proper estimation of the relative frequency of problematic cases, it is unclear whether it is even possible to reject empirical adaptationism by accumulating problematic cases.

In sum, I have used three problematic cases to show that it is not always possible to compare the relative causal importance of natural selection and non-selective evolutionary factors in an intelligible way. Each type of these problematic cases has multiple instances across a broad diversity of species or higher taxa, and these problematic cases are not intended to be exhaustive. The widespread existence of problematic cases provides an important reason to be skeptical of the testability of claim (I).

4. Orzack and Sober on empirical adaptationism

I now turn to claim (II): Natural selection is in most cases the *only important* cause of phenotypic evolution. A formulation of empirical adaptationism like this has been proposed by Orzack and

Sober. Orzack and Sober (1994a, p. 362) distinguish three propositions about the role of natural selection in the evolution of some individual trait *T* in a given population:

> (U) Natural selection played some role in the evolution of *T*.
> (I) Natural selection was an important cause of the evolution of *T*.
> (O) Natural selection is a sufficient explanation of the evolution of *T*, and *T* is locally optimal.

Then they formulate the thesis of empirical adaptationism as a generalized form of proposition (O):

> Natural selection is a sufficient explanation for most nonmolecular traits, and these traits are locally optimal. (Orzack and Sober 1994a, p. 364)

How could biologists tell whether natural selection is a sufficient explanation of the evolution of a trait? Orzack and Sober suggest that this can be tested via an optimality model. An optimality model is a kind of "censored" model in which non-selective evolutionary processes such as genetic drift are ignored and only natural selection is described. Under this setting, the most adaptive variant of a trait among a set of alternatives in a particular environment is expected to be selected and eventually driven to fixation in the population. If the optimality model's quantitative prediction about the optimal value of an individual trait in a given population fits the empirical data statistically, and if there is no significant between-individual heterogeneity with respect to this fit, then natural selection would be a sufficient explanation of the evolution of that trait. Under these circumstances, natural selection is also said to be the *only important* cause of the evolution of the focal trait. Hence, in another place Sober (1998, p. 72) also formulates empirical adaptationism as follows:

> Natural selection has been the only important cause of most of the phenotypic traits found in most species.

For Orzack and Sober, claiming that natural selection is the only important cause of the evolution of a trait does not mean that it is the only cause at work. In these cases, non-selective

evolutionary factors may still exist, but they play such a small role in the evolution of a trait that they can be ignored without loss of accuracy when predicting or explaining the evolutionary outcome of that trait. For example, since no real biological population is infinitely big, genetic drift always plays some role in the evolution of a population. But when the effective size of a population is big enough, it is usually innocuous to ignore the effect of genetic drift when studying the evolution of a trait in that population.

It is worth emphasizing again that Orzack and Sober's formulation of empirical adaptationism is different from claim (I) that natural selection is in most cases the most important cause of phenotypic evolution. After all, it is possible for natural selection to be the most important, but not the only important, cause of the evolution of a trait. This subtle but important difference between Orzack and Sober's formulation and claim (I) has failed to be fully appreciated in the literature. For example, Resnik (1997, p. 42) describes Sober's definition of empirical adaptationism as the claim that "natural selection is the most important cause of most traits in most populations"; Lewens (2009, p. 163) summarizes Orzack and Sober's formulation of empirical adaptationism as the view that " selection is typically the most important evolutionary force." Both conflate Orzack and Sober's formulation, which is a version of claim (II), with claim (I).

One major benefit of formulating empirical adaptationism in Orzack and Sober's way is that it does not involve *directly comparing* the relative causal importance of natural selection and non-selective evolutionary factors, thus circumventing the comparison problem faced by claim (I). This is because Orzack and Sober's test is about whether natural selection is a sufficient explanation of the evolution of a phenotypic trait. This test requires some kind of comparison,

but such comparison is made between empirical data and the predictions of an optimality model, not between the causal importance of natural selection and non-selective evolutionary factors.

One may worry that I overemphasize the difference between claims (I) and (II). If natural selection is in most cases the *only important* cause of phenotypic evolution, it should follow that natural selection is in most cases the *most important* cause of phenotypic evolution. In other words, claim (II) entails claim (I). Given this, it seems that those who want to test claim (I) can circumvent the comparison problem by using Orzack and Sober's strategy to first test claim (II). If claim (II) is true, then claim (I) is also true.

Despite its appeal, this strategy has a serious limitation: Although the truth of claim (II) entails the truth of claim (I), one cannot assure the truth of claim (II) before testing. If claim (II) is tested to be false, one would not be able to tell whether claim (I) is true or false. In other words, the testing method of claim (II), such as the strategy used by Orzack and Sober, does not provide a *full* test of claim (I). Hence, although claim (II) entails claim (I), it is still important to keep them distinct.

5. Two methodological difficulties in the long-run test of empirical adaptationism

In this section, I will show that Orzack and Sober's suggested test of empirical adaptationism is faced with two serious issues: The first concerns the difficulty of meaningfully aggregating testing results of optimality models with varying backward time scales; the second concerns the difficulty of obtaining a representative sample of phenotypic traits due to the poorly defined statistical population. These issues, if cannot be resolved, will undermine the status of empirical adaptationism as a meaningful empirical thesis that is testable in a biologically non-arbitrary way.

5.1 The difficulty of meaningfully aggregating testing results of optimality models with varying backward time scales

In a comment on Orzack and Sober's (1994) suggested test of empirical adaptationism, Brandon and Rausher (1996) point out that Orzack and Sober fail to indicate the intended backward time scales of their optimality models. Whether natural selection can be said to provide a sufficient explanation of the evolution of a trait partly depends on how far back an evolutionary biologist intends her optimality model to extend. For the same trait, natural selection may be a sufficient explanation of its evolution over a short period of time in the past but fails to be so if we go back far enough in time.

To illustrate this point, consider again Case 3 discussed in Section 3. A population consists of three variants (V_1, V_2, V_3) of a trait T, in which Fitness (V_1) > Fitness (V_2) > Fitness (V_3). V_1 is rare and ultimately lost from the population because of genetic drift. Given the absence of V_1, V_2 is favored by natural selection and driven to fixation. If we start considering the evolution of T from the time point at which V_1 disappeared, then natural selection can be said to provide a sufficient explanation of its evolution. However, if we trace back to the time when all three variants are present, then natural selection cannot be the only important cause of the evolution of T, because genetic drift is also needed to explain why V_2, instead of V_1, ultimately gets fixed in the population.

Although critical of the ambiguity concerning the backward time scale in Orzack and Sober's approach, Brandon and Rausher do not take it to be a fatal issue. Instead, they think that such an ambiguity can be resolved by specifying that whether natural selection is the sole process involved in the evolution of a trait should be assessed "from the point in time at which all of the relevant variants exist in the relevant lineage" (Brandon and Rausher 1996, p. 192).

I agree with Brandon and Rausher's suggestion when it comes to constructing optimality models of *specific* phenotypic traits. But at the same time, I think that adopting their resolution method will lead to another problem for the test of empirical adaptationism. Empirical adaptationism concerns the overall relative causal importance of natural selection in evolution, which must be evaluated *relative to a specific time period*. However, for different traits of different species, the time points at which all the relevant variants exist in the relevant lineages can be dramatically different. For example, evolutionary biologist Richard Lenski and his students started a long-term evolution experiment in 1988 with twelve genetically identical populations of *E. coli*. In 2003 (around generation 31500), they found that some bacteria in one population had evolved the ability to feed on citrate in the presence of oxygen and quickly dominated that population (Blount et al. 2008). If researchers want to use an optimality model to study the role of natural selection in the evolution of this new trait, they only need to trace back to the year 2003 according to Brandon and Rausher's suggestion. Things are rather different for researchers who study the evolution of the West Indian Ocean coelacanth. This kind of fish is usually considered as a "living fossil" because it shows little morphological divergence compared with the coelacanth fossils formed approximately 400 million years ago. Living at a depth of about 200 m, coelacanths have evolved the ability to detect a narrow range of light at about 480 nm. According to Yokoyama and Tada's (2000) estimation, the adaptation of coelacanths towards the deep-sea started as early as 200 million years ago. Following Brandon and Rausher's suggestion, the optimality model used to study the evolution of coelacanths' specialized light detection ability needs to cover a time scale of about 200 million years.

Orzack and Sober (1994a, p. 365) suggest that empirical adaptationism is testable "by the accumulation of successes and failures of specific optimality models." In the case of *E. coli*, the

optimality model is used to evaluate the role of natural selection over a time scale no more than two decades. In the case of the West Indian Ocean coelacanth, however, the optimality model evaluates the role of natural selection over a time scale as long as 200 million years. There is no doubt that the testing result of each optimality model provides some information about the causal importance of natural selection in an evolutionary event that occurred during a certain time period. The difficulty lies in how to determine the *weight* of such causal information provided by each test. According to Orzack and Sober's counting rule, each testing result has the same weight: For any optimality model, each success or failure counts one in the ensemble test, regardless of the length of time scale it covers. However, it is clear that the weight of causal information provided by a case where natural selection is shown to be the only important cause of phenotypic evolution during 200 million years is *not the same* as that provided by a case where natural selection is the only important cause of evolution during merely two decades. Hence, simply aggregating successes and failures of optimality models with dramatically different backward time scales does not have a straightforward biological meaning as Orzack and Sober have assumed; it does not provide a proper evaluation of the overall relative causal importance of natural selection in evolution relative to a specific time period.

Proponents of Orzack and Sober's approach may argue that one can assign different weights to the testing results of optimality models with different backward time scales. For example, if natural selection is shown to the only important cause of the evolution of a trait during a one-million-year interval, this testing result counts one million "positive" points; if not, it counts one million "negative" points. However, this strategy does not work, because it assumes that the causal power of natural selection is distributed homogeneously in time, which is something we cannot know through the testing of optimality models.

It seems that researchers who want to test empirical adaptationism by using optimality models are faced with a dilemma. If they choose to follow Brandon and Rausher's suggestion, then they are faced with the difficulty of meaningfully aggregating testing results of optimality models with varying backward time scales. If instead they choose a fixed time scale and use it for the construction of every optimality model, then they can avoid the aggregation problem above, but will be faced with two other issues. First, the choice of this time scale will be unavoidably arbitrary. Second, no matter how long or short this time scale is, it cannot fit the evaluation of *every* trait. The chosen time scale may be too long for evaluating the evolution of some traits, because when we go back too far in time, the relevant variants (and even the relevant species) may not have yet existed. It may be, however, too short for evaluating the evolution of other traits, for the relevant variants may have appeared in the lineage long before the chosen time period.

5.2 The difficulty of obtaining a representative sample of phenotypic traits due to the poorly defined statistical population

Empirical adaptationism as formulated by Orzack and Sober is a *general* thesis about nature – a claim about the overall relative causal importance of natural selection in evolution. As an implication, the correctness of such a thesis cannot be assessed by testing the local optimality of a single phenotypic trait. Orzack and Sober are fully aware of the ensemble nature of such a test, and they suggest that empirical adaptationism is testable by accumulating the testing results of many traits:

> The test of adaptationism we advocate need not engender an interminable debate. Forty or 50 appropriately structured studies might well provide a reasonable assessment of adaptationism. For example, if 45 of the 50 tests lead to the conclusion that the trait in question is locally optimal, in our opinion one

could conclude that adaptationism is correct. Attainment of some agreed-on
number of tests should be a goal of evolutionary biologists […] [A] test of
adaptationism of the size mentioned earlier might even be attainable in the next
10 yr or so. (Orzack and Sober 1994a, pp. 377–378)

28 years have passed since Orzack and Sober made this proposal, but the kind of test envisaged

by them has yet to be conducted. But even if biologists decided to implement such a project, they

would have to first address a number of methodological difficulties to guarantee the validity of

the test.

In an ideal case, biologists would have a complete list of all the phenotypic traits of all

organisms in the history of life, and they are able to test in each case whether natural selection is

sufficient to explain the evolution of a trait. Clearly, this is not a practical plan. A more feasible

way to conduct the test is to choose a *representative* sample of phenotypic traits.

Several methodological issues need to be addressed in order to obtain such a

representative sample. First, evolutionary biologists need to reach some kind of consensus on the

number of traits to be studied in the sample. Orzack and Sober (1994a, p. 377) suggest that

"[f]orty or 50 appropriately structured studies might well provide a reasonable assessment of

adaptationism." But they provide no clue of how this number is determined and justified from a

statistical point of view. Even if the issue of sample size can be resolved, another methodological

difficulty remains, namely, how to select a sample of traits such that the testing result is

representative of the overall relative causal importance of natural selection in the evolution of all

the traits of all organisms appearing in the history of life.

It may be argued that statisticians have developed various sampling techniques to help

ensure the representativeness of a sample. For example, if there are two colors of glass beads –

red and blue – in a box and a researcher wants to know the proportion of the red ones, she can

mix the beads sufficiently and select a sample of a certain size randomly. This sampling

technique is called *simple random sampling*, and the underlying rationale is that in such a setting, every possible sample of the same size has the same probability of being selected during sampling.

However, sampling methods like this may not be easily applied to the case of testing empirical adaptationism. One major goal of statistical research is to draw conclusions about the properties of populations via studying samples. By "population," or more precisely, "statistical population," I mean a group of individuals that a researcher wants to draw conclusions about, which can be either concrete objects such as glass beads, or abstract objects such as the possible moves of a chess player. From a statistical point of view, having a *well-defined* statistical population is essential for getting a representative sample of that population. A statistical population is well-defined if and only if there is a clear standard about what should be included in the population and each individual in the population is possible to be sampled during the time of investigation. In the case of glass beads, the statistical population is the collection of all the glass beads in the box, which is well-defined according to the preceding definition. Without the existence of a well-defined statistical population, it does not even make sense to ask whether a sample is representative, because the representativeness of a sample, by definition, is always relative to the statistical population it is intended to provide information about.

Now consider the case of testing empirical adaptationism. What is the statistical population that biologists want to know about? A quick answer would be "all the traits of all organisms." But this answer is too vague to be helpful. For example, should structures like eyes or wings of different species be regarded as one trait? Or should each of them count as a different trait? A clue of answers to these questions may be found in the subtle changes in the way empirical adaptationism is formulated by Sober in different places:

(i) Natural selection is a sufficient explanation for *most nonmolecular traits*, and these traits are locally optimal (Orzack and Sober 1994a, p. 364, my italics).

(ii) Natural selection has been the only important cause of *most of the phenotypic traits found in most species* (Sober 1998, p. 72, my italics).

(iii) *Most phenotypic traits in most populations* can be explained by a model in which selection is described and nonselective processes are ignored (Sober 2000, p. 124, my italics).

Formulation (i) talks of nonmolecular traits (i.e., phenotypic traits) in general. Formulation (ii) seems to suggest that analyses of phenotypic traits should be at the level of species. Formulation (iii) goes one step further: Studies of the same trait in *different populations* of a species may yield different results. For example, a trait that is locally optimal in one population may not be so in another. Hence, it seems reasonable to regard studies of the same trait in different populations as different individual cases while sampling. Suppose that we adopt the granularity of analysis in formulation (iii), then the statistical population being studied would be all the phenotypic traits in each population of each species.

However, such a statistical population is still poorly defined. Although the concept of trait is widely used in biology, it is notoriously difficult to define what a trait is. In a very general sense, a trait is simply a character state of an organism. But exactly what feature of an organism could be properly regarded as a trait depends on many factors, including facts of developmental processes of organisms (Wagner 2001), researchers' background theories and beliefs (Resnik 1997), and the specific content of the research problem in question. Hence, different biologists would identify different collections of traits even within one population of one species, and the situation will be even worse when considering all the populations of all species. But the problem goes beyond this.

In the case of glass beads, each glass bead in the box has the same probability of being selected in the process of sampling. In the case of testing empirical adaptationism, however, many features of organisms have no chance to be sampled because they have yet to be individualized by biologists as traits and hence cannot enter the sampling pool at all. In other words, the statistical population intended to include all the phenotypic traits in each population of each species is not well-defined, which makes it impossible to collect a representative sample thereof.

It may be argued that instead of trying to obtain a representative sample of all the phenotypic traits in each population of each species, biologists can just build a list of all the traits that have already been individualized and studied. If most already-studied traits are locally optimal, then Orzack and Sober's version of empirical adaptationism is true. This approach does not really solve the problem of representativeness. The way we distribute our scientific resources can influence our perception of the relative importance of natural selection in evolution (Beatty 1987, pp. 53–54). The more scientific resources are distributed to studies on the apparent design of organisms and their relations of adaptedness to their environments, the more likely it is to find cases of local optimality. Similarly, the more funding is provided to studies on the role of non-selective evolutionary factors (such as genetic drift) in evolution, the more likely it is that counterexamples of optimality can be found. If the testing result of empirical adaptationism is simply a statistical summary of the results of biologists' previous studies, then this result would be more of a reflection of scientific resource distribution and biologists' research interest rather than what actually happens in nature. This observation poses a serious problem for those in favor of testing empirical adaptationism, for it contradicts with one of their core beliefs that empirical adaptationism is an empirical thesis about *nature*.

5.3 Replies to possible objections

Now I will consider some possible objections to my above discussion. The first is that my critiques only target Orzack and Sober's approach, which have no implications for the testability of empirical adaptationism formulated in *other* ways.

Although I use Orzack and Sober's proposal as a target of critique, the two difficulties I have identified present *general* challenges for any serious attempt to test empirical adaptationism by accumulating testing results of particular phenotypic traits. First, the evaluation of the causal role of natural selection in the evolution of any phenotypic trait must be conducted relative to a specific time scale, regardless of what kind of models or strategies are used to perform such evaluation. As I have shown in Section 5.1, it is virtually impossible to find one time scale that can fit the evaluation of all phenotypic traits. Insofar as the correctness of empirical adaptationism is assessed via an ensemble test, researchers will be faced with the difficulty of meaningfully aggregating testing results of different phenotypic traits relative to varying time scales. Second, assessing the overall relative causal importance of natural selection inevitably involves testing a sample of phenotypic traits. Hence, anyone who is serious about testing empirical adaptationism has to deal with the difficulty of obtaining a representative sample of all phenotypic traits.

The second possible objection is that the difficulties I have identified are *merely practical* or *pragmatic*, which have no implications for the testability of empirical adaptationism *in principle*. To address this objection, first we need to clarify what it means when a difficulty is described as "merely practical." Here is what I take to be a reasonable definition: A difficulty is merely practical if and only if it is possible to be resolved with the advance of science and

technology and/or the investment of more resources (time, money, labor, etc.). Given this definition, the problem of aggregating testing results of different phenotypic traits relative to varying time scales is not merely practical, because the difficulty arises from the very nature of the relevant research objects: Different phenotypic traits of different species evolve at different rates and are most properly evaluated relative to different time scales; this fact will not be changed by the advance of science and technology or the investment of more resources.

As for the difficulty of obtaining a representative sample, one may argue that it is a challenge for statistical research in general, not unique to the test of empirical adaptationism. Also, it is merely a practical concern because the representativity of samples comes in degrees and it can be improved by developing better sampling methods or increasing the size of sample. This objection makes sense in cases where the statistical population under investigation is well defined. In the most extreme case, researchers can study a statistical population of interest by examining every individual in it rather than by choosing a sample. But even this method does not apply to the case of testing empirical adaptationism: Since the statistical population of interest – the population consisting of all the phenotypic traits in each population of each species – is poorly defined, biologists do not even have a common population as their object of study, making it impossible to examine every individual in that population or obtain a representative sample thereof. This problem cannot be dismissed as "merely practical."

The third objection is that the worry about having a poorly defined statistical population of phenotypic traits presents a challenge for making any generalization about traits, not about empirical adaptationism in particular. If my criticism of empirical adaptationism is accepted, it would yield the disastrous conclusion that generalizing about phenotypic traits is hopeless in biology. To address this objection, it is important to realize that there are different kinds of

generalizations in the biological sciences, which may have different forms and bases (Cooper 1998). The challenges faced by empirical adaptationism are not necessarily faced by other kinds of generalizations about phenotypic traits.

To elaborate this point, let us first consider why having a poorly defined statistical population is a particularly acute problem for empirical adaptationism. The kind of generalization involved in empirical adaptationism is essentially a statistical inference about a population based on data collected on a sample of that population. Indeed, there is no deeper principle or mechanism explaining why we should anticipate the correctness of empirical adaptationism. Hence, the only way to test its validity is to have a reasonably representative sample, which requires a reasonably well-defined statistical population. Of course, the well-definedness of a statistical population can come in degrees. It goes beyond the scope of this article to provide a systematic account of how to quantitatively measure the well-definedness of a statistical population. But no matter how well-definedness turns out to be measured, the supposed statistical population in the case of empirical adaptationism – the population consisting of all the phenotypic traits in each population of each species – is bound to have a very low degree of well-definedness, given the reasons explicated in Section 5.2. And it is unclear how this low degree of well-definedness can be significantly improved in the case of empirical adaptationism.

However, not all generalizations about traits in biology are purely statistical inferences, nor do all generalizations concern "all the phenotypic traits in each population of each species." For example, Allen's rule says that warm-blooded animals tend to have shorter protruding body parts (e.g., bills, tails, ears) in cooler climates (Mayr 1963, pp. 320–324). For many biologists, this generalization is not a purely statistical inference based on a sample, but something that

could be anticipated theoretically, for shorter protruding body parts can help reduce heat loss in cooler climates, which provides a selective advantage. By contrast to empirical adaptationism, Allen's rule does not concern the totality of phenotypic traits but certain specific traits. In addition, since there are numerous exceptions, empirical studies on Allen's rule are not so much to test whether this rule holds in more than 50% of cases as to find the circumstances under which it does and does not hold. Hence, the worry about having a poorly defined statistical population of all phenotypic traits in the case of empirical adaptationism does not exclude the possibility of making other types of generalizations about traits such as Allen's rule.

6. Rethinking the value and necessity of testing empirical adaptationism

The previous sections have focused on the testability of empirical adaptationism. I now turn to a different issue: Why should biologists care about the truth of empirical adaptationism at all? As mentioned earlier, a number of philosophers and some biologists have tried to clarify the meaning of empirical adaptationism and provide their own formulations. These scholars are not necessarily committed to the truth of empirical adaptationism; they leave the task of testing it to biologists. When doing so, many if not all of them also assume, explicitly or implicitly, that empirical adaptationism is worth testing.[3] However, in the existing literature, there has been little explicit justification of why this should be the case.

The earlier sections discussing the testability of empirical adaptationism already provide one reason to be skeptical of the values of testing this thesis: Given that empirical adaptationism

---

[3] Orzack and Sober (1996) explicitly claim that testing empirical adaptationism is scientifically valuable. Godfrey-Smith (2001) devotes some serious discussion to how empirical adaptationism should be tested, which makes it reasonable to believe that he at least *implicitly* assumes such a test to be valuable. Lewens (2009) suggests some ways to clarify the meaning of adaptationism understood as an empirical hypothesis. When doing so, he may not really consider whether such a hypothesis is worth testing, but mainly aims to clarify the confusion in debates over adaptationism.

is not as well-formulated and testable as one might think, any defense of the values of testing this thesis would fail insofar as these alleged values are based on the actual testing result. Nevertheless, it is still important to give a more direct critical analysis of the value and necessity of testing empirical adaptationism in case some readers are not convinced by my argument against its testability.

In this section, I will consider four types of value that might be attached to testing empirical adaptationism. It turns out that none of them can be used to justify the necessity of conducting such a test. To be clear, the four types of value I will consider below are not intended to be exhaustive of all conceivable values. Hence, my analysis here does not allow me to conclude that no other values of testing empirical adaptationism can be found in the future. Nevertheless, my argument, if succeeds, will show that those in favor of testing empirical adaptationism have yet to demonstrate the distinctive value and necessity of conducting such a test. Unless a proper justification can be provided, there seems no reason to assume that empirical adaptationism is worth testing.

6.1 Methodological heuristic value

As mentioned in the introduction, there are different kinds of adaptationism. Those in favor of testing empirical adaptationism may seek to argue for the necessity of such a test by appealing to the relationship between empirical adaptationism and another kind of adaptationist position – methodological adaptationism. Godfrey-Smith (2001, p. 337) defines methodological adaptationism as the following view: "The best way for scientists to approach biological systems is to look for features of adaptation and good design. Adaptation is a good 'organizing concept' for evolutionary research." Lloyd (2015), a representative critic of methodological adaptationism,

formulates it in terms of the research question asked by its practitioners: A methodological adaptationist assumes, at the beginning of investigation, that a trait under consideration is an adaptation and asks "What is the function of this trait?"; given this research question, they would try to look for adaptative explanations for the evolution of this trait.

Those in favor of testing empirical adaptationism may argue that many biologists are motivated to adopt methodological adaptationism in actual scientific research because they believe the truth of empirical adaptationism: The belief in the power and ubiquity of natural selection in evolution motivates them to assume, at the beginning of their investigation, that a trait is an adaptation until shown otherwise. For these researchers, it truly matters whether empirical adaptationism is correct, because methodological adaptationism – a widely-practiced methodology – is justified by the truth of empirical adaptationism. Given the central role of empirical adaptationism in motivating and justifying actual scientific practice, it is valuable and necessary to test it.

This argument does not work because it fails to distinguish the *beliefs* motivating someone to adopt a methodology and the *genuine justification* needed to defend such a methodology. It is perfectly possible that a methodological adaptationist is motivated to adopt this methodology because she *believes* that empirical adaptationism is true and that the truth of empirical adaptationism justifies the legitimacy of methodological adaptationism as a valid research strategy. However, this does not mean that the justification of methodological adaptationism *actually* relies on the truth of empirical adaptationism. Godfrey-Smith (2001, p. 338) has shown that empirical adaptationism and methodological adaptationism are logically independent of each other. Therefore, the truth or falsity of empirical adaptationism does not really bear on the justifiability of methodological adaptationism – even if empirical

adaptationism turns out to be false, it does not exclude the possibility of methodological adaptationism being a valid and productive research strategy. Notice that I am not denying the fact that empirical adaptationist beliefs motivate some (or even many) biologists to adopt methodological adaptationism. But this fact does not guarantee the value and necessity of testing empirical adaptationism. To evaluate the validity of methodological adaptationism, we need to examine its merits (or problems) from a methodological perspective, rather than test the truth of empirical adaptationism.

This analysis is supported by the actual critiques offered by opponents of methodological adaptationism. For example, in her systematic critique of methodological adaptationism, Lloyd (2015) objects to this research method by identifying various dangers that result from the logic of its research question, such as the lack of a stopping rule in pursuing adaptive explanations for the evolution of traits and the loss of ability to evaluate and weigh evidence for alternative causal hypotheses. None of these critiques is based on the falsity of empirical adaptationism. Lloyd also proposes an alternative research strategy called the "evolutionary factors framework" and notes that "the evolutionary factors framework is independent of any commitment regarding empirical (or 'metaphysical') adaptationism" (Lloyd 2015, p. 345). It goes beyond the scope of this article to evaluate whether the evolutionary factors framework is indeed a better alternative to methodological adaptationism. What is relevant here is that, for Lloyd, the reason why biologists should abandon methodological adaptationism and adopt the evolutionary factors framework has nothing to do with the truth or falsity of empirical adaptationism, but with the methodological superiority of the evolutionary factors framework in biological research.

6.2 Explanatory value

It may be argued that empirical adaptationism is worth testing because its truth (or falsity) has explanatory value. A piece of information has explanatory value in science if and only if it contributes to explaining certain scientific phenomena. Now the question is: Can we find any biological phenomenon whose explanation at least partly relies on the testing result of empirical adaptationism? It is easy to give some superficial examples such as "The truth of empirical adaptationism helps explain the evolution of complex adaptive traits." However, a closer look at such examples reveals that what does the explanatory work is the relevant evolutionary factors such as natural selection, not the correctness of empirical adaptationism. In my view, prospects are dim for finding genuine examples where the truth or falsity of empirical adaptationism has explanatory power. I have presented some serious difficulties of testing empirical adaptationism in previous sections. But even if it turns out to be testable, its truth (or falsity) is simply based on a statistical summary of the testing results of individual cases; it is unclear how such a highly contingent result can help explain the occurrence of any biological phenomenon.

One possible objection is that while the testing result of empirical adaptationism per se has no explanatory value, testing this thesis requires testing specific hypotheses about the causal role of natural selection in the evolution of particular traits. The results of such tests would contribute to explaining the form and distribution of those particular traits. This argument is problematic because it seems to assume that testing empirical adaptationism is a *necessary* condition for testing hypotheses about the evolution of particular traits when this is actually not the case. Since the truth (or falsity) of empirical adaptationism has no bearing on why a particular trait has evolved, biologists do not need to test empirical adaptationism in order to pursue and test hypotheses about the evolution of particular traits.

6.3 Epistemic value

Resnik (1997) argues that Sober and Orzack's long-run test of empirical adaptationism is valuable because it can "increase our knowledge about evolutionary trends" (Resnik 1997, p. 46). If testing empirical adaptationism can increase our biological knowledge, then it seems to have epistemic value.

There are two issues with this proposal. First, the testing result of empirical adaptationism does not reveal an evolutionary trend. Evolutionary trends are directional changes over long periods of time. The testing result of empirical adaptationism at best tells us the overall relative causal importance of natural selection in phenotypic evolution relative to a specific period of time, but it tells nothing about how the overall relative causal importance of natural selection will change in the future. If this is the case, then it leads to the second issue, that is, it remains unexplained *why* it is scientifically valuable to know the testing result of empirical adaptationism. Failing to answer this question is tantamount to saying that it is scientifically valuable to test empirical adaptationism because it is scientifically valuable to know the result of such a test, which merely begs the question.

It may be argued that testing empirical adaptationism is valuable for those who want to conduct such a test because it satisfies their curiosity, and there is no need to ask further what the value of such curiosity is. Notice that this justification focuses only on the value of testing empirical adaptationism in terms of satisfying *individual curiosity*, but it says nothing about the value of such a test to science in general and biology in particular. An inquiry that arouses the curiosity of certain individuals may not be of scientific value. For example, an individual might want to know the center of gravity of all the socks on Earth, but knowledge of such a fact does not seem to have any scientific value. By the same token, testing empirical adaptationism may be

interesting to some researchers, but this does not automatically demonstrate its *scientific value*. If we view scientific research as a social practice and scientific knowledge as social knowledge (Longino 1990), then individual scientists need to provide further justification about why certain questions of interest to themselves are scientifically valuable and hence worth investigating and discussing within the scientific community.

6.4 Spin-off value

Not all scientific controversies can be resolved in the end. In many cases, participants in a scientific debate simply lose their interest and move to other topics (Kovaka 2021). Nevertheless, these debates usually promote many meaningful discussions along the way, creating great spin-off value. In the case of empirical adaptationism, one may concede that there is no conclusive answer with respect to its correctness, but still contend that the debates surrounding empirical adaptationism have inspired many in-depth discussions about a number of important topics in evolutionary biology. Given this, it seems reasonable to say that testing empirical adaptationism has some spin-off value.

Historically speaking, there is some element of truth in the above claims. For example, given Orzack and Sober's formulation of empirical adaptationism, evaluating the correctness of this thesis inevitably involves assessing the validity of the optimality approach in evolutionary research. This awareness has motivated them to publish a series of papers with the aim of clarifying how to properly construct and test an optimality model (Orzack and Sober 1994a; 1994b; 1996), which are in no doubt scientifically valuable for evolutionary research. But facts like these do not justify the necessity of *continuing* the debates over empirical adaptationism or trying to test its correctness. Potochnik (2009) has convincingly shown that the fate of optimality

modeling is not necessarily linked to that of empirical adaptationism. No matter whether empirical adaptationism is testable or not, and if it is testable, no matter whether it is true, the centrality of various uses of optimality models ensures a continuing role for the optimality approach in evolutionary research. Hence, stopping the debate about the truth of empirical adaptationism will not hinder the study of optimality modeling. In fact, much confusion can be avoided if studies like this can be detached from the debates over empirical adaptationism.

7. Conclusion: Rethinking the scientific debates involving overall relative causal importance

While empirical adaptationism is typically described as an empirical claim about nature, few serious efforts have been made by biologists to actually test this general claim. In some sense, the debates over empirical adaptationism are no longer a live issue in biology or the philosophy of biology. If so, why do we need another article on this topic?

Two observations have motivated me to rethink empirical adaptationism. First, the mainstream literature on empirical adaptationism still regards it as an empirical claim that can be and should be tested, which has to some extent become a dogma in the philosophy of biology. Second, while few biologists have actually tried to test empirical adaptationism, they sometimes talk as if it were true (e.g., Losos 2014).

However, as I have tried to argue in this article, there exists a series of conceptual and methodological difficulties that may well undermine the testability of empirical adaptationism. Also, those in favor of testing empirical adaptationism have yet to demonstrate the distinctive value and necessity of conducing such a test. If this diagnosis is correct, it may partially explain why few currently-active biologists have actually tried to test empirical adaptationism. Rather

than attempt to reignite debates over empirical adaptationism, I suggest that it is probably time to say goodbye to it and move on. Instead of asking "Which evolutionary factor is *generally* more important?", it may be more productive to ask "How do different evolutionary factors interact with each other to influence the course of evolution?".

The case of empirical adaptationism also provides reasons for scientists to reconsider the value and necessity of engaging in debates involving the notion of overall relative causal importance. Such debates are actually quite common in scientific research. For example, in the nature-nurture debate, researchers argue about whether genes or environmental factors generally play a more important role in human development; in the niche-neutral debate, ecologists disagree about whether niche-based processes or neutral processes are generally more important in structuring ecological communities (Chase 2014); in the field of cultural evolution, there is the debate about whether cultural transmission is generally more influenced by preservative processes or transformative processes (Acerbi and Mesoudi 2015). All these debates concern the overall relative causal importance of different factors with respect to the totality of phenomena in a domain, and participants in these debates usually take the value and necessity of engaging in such debates for granted. However, my analysis of the case of empirical adaptationism has shown that this kind of "taking-for-granted" can be very problematic. On the one hand, in many cases it is just impossible to compare the relative causal importance of different factors in an intelligible way. On the other hand, even when this kind of comparison is possible, the sheer generality of debates involving the notion of overall relative causal importance may leave us unclear about the value of such debates, especially given the fact that the relative causal importance of different factors in a domain can be very context-dependent and may change from case to case. The applicability of these problems to various debates should be assessed case by

case, but the general lesson is that scientists should change their *default* attitude towards such debates. Instead of assuming that debates involving the overall relative causal importance of different factors are self-evidently necessary and valuable, scientists should evaluate, case by case, the value and necessity of engaging in such debates. When an alleged empirical debate involves claims that are confusingly vague, or when it is not clear what the scientific value of such a debate is, it is better to stop and think about whether it is worthwhile to step into this debate and whether there is a more productive way to structure the discussion.

References

Acerbi A, Mesoudi A (2015) If we are all cultural Darwinians what's the fuss about? Clarifying recent disagreements in the field of cultural evolution. Biol Philos 30:481–503.

Austad SN, Hoffman JM (2018) Is antagonistic pleiotropy ubiquitous in aging biology? Evol Med Public Heal 2018:287–294

Beatty J (1997) Why Do Biologists Argue like They Do? Philos Sci 64:S432-S443

Beatty J (1987) Natural selection and the null hypothesis. In: Dupré J (ed) The Latest on the Best: Essays on Evolution and Optimality. MIT Press, Cambridge, MA and London, England, pp 53–75

Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key

    innovation in an experimental population of *Escherichia coli*. Proc Natl Acad Sci

    105:7899–7906

Brandon RN, Rausher MD (1996) Testing Adaptationism: A Comment on Orzack and Sober.

    Am Nat 148:189–201

Chase JM (2014) Spatial scale resolves the niche versus neutral theory debate. J Veg Sci 25:319–

    322.

Chouteau M, Angers B (2012) Wright's shifting balance theory and the diversification of

    aposematic signals. PLoS One 7:e34028.

Cooper G (1998) Generalizations in ecology: A philosophical taxonomy. Biol Philos 13:555–586.

Darwin C (1859) On the Origin of Species, 1st edn. John Murray, London

Darwin C (1869) On the Origin of Species, 5th edn. John Murray, London

Godfrey-Smith P (2001) Three Kinds of Adaptationism. In: Orzack SH, Sober E (eds)

    Adaptationism and Optimality. Cambridge University Press, Cambridge, pp 335–357

Gould SJ, Lewontin RC (1979) The Spandrels of San Marco and the Panglossian Paradigm: A

    Critique of the Adaptationist Programme. Proc R Soc London 205:581–598

Haldane JBS (1927) A Mathematical Theory of Natural and Artificial Selection, Part V:

    Selection and Mutation. Math Proc Cambridge Philos Soc 23:838–844

Kimura M, Ohta T (1971) Theoretical Aspects of Population Genetics. Princeton University

    Press, Princeton, New Jersey

Kovaka K (2021) Fighting about frequency. Synthese 199:7777-7797

Lewens T (2009) Seven types of adaptationism. Biol Philos 24:161–182

Lloyd EA (2015) Adaptationism and the Logic of Research Questions: How to Think Clearly About Evolutionary Causes. Biol Theory 10:343–362.

Longino HE (1990) Science as social knowledge: Values and objectivity in scientific inquiry. Princeton University Press, Princeton, NJ

Losos JB (2014) What Is Evolution? In: Losos JB, Baum DA, Futuyma DJ, et al. (eds) The Princeton Guide to Evolution. Princeton University Press, Princeton and Oxford

Mayr E (1983) How to Carry Out the Adaptationist Program? Am Nat 121:324–334

Mayr E (1963) Animal Species and Evolution. Harvard University Press, Cambridge

Millstein RL (2008) Distinguishing drift and selection empirically: "The great snail debate" of the 1950s. J Hist Biol 41:339–367

Orzack SH, Forber P (2010) Adaptationism. Stanford Encycl. Philos.

Orzack SH, Sober E (1994a) Optimality Models and the Test of Adaptationism. Am Nat 143:361–380

Orzack SH, Sober E (1994b) How (not) to test an optimality model. Trends Ecol Evol 9:265–267.

Orzack SH, Sober E (1996) How to Formulate and Test Adaptationism. Am Nat 148:202–210

Potochnik A (2009) Optimality modeling in a suboptimal world. Biol Philos 24:183–197.

Resnik D (1997) Adaptationism: Hypothesis or Heuristic. Biol Philos 12:39–50

Skipper RA (2002) The persistence of the R.A. Fisher-Sewall Wright controversy. Biol Philos 17:341–367.

Sober E (1987) What Is Adaptationism? In: Dupré J (ed) The Latest on the Best: Essays on Evolution and Optimality. MIT Press, Cambridge, pp 105–118

Sober E (1998) Six Sayings about Adaptationism. In: Hull DL, Ruse M (eds) The Philosophy of Biology. Oxford University Press, New York, pp 72–86

Sober E (2000) Adaptationism. In: Philosophy of Biology, 2nd edn. Westview Press, Boulder, CO, pp 121–145

Wade MJ, Goodnight CJ (1991) Wright's Shifting Balance Theory: An Experimental Study. Science 253:1015–1018

Wade MJ, Goodnight CJ (1998) Perspective: the theories of Fisher and Wright in the context of metapopulations: when nature does many small experiments. Evolution 52:1537–1553.

Wagner GP (2001) The character concept in evolutionary biology. Academic Press, San Diego

Wright S (1931) Evolution in Mendelian populations. Genetics 16:97–159

Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: Proceedings of the Sixth International Congress of Genetics. pp 356–366

Yokoyama S, Tada T (2000) Adaptive evolution of the African and Indonesian coelacanths to deep-sea environments. Gene 261:35–42