Forthcoming in *The Journal of Philosophy*. Please quote only the published version.

MUCH ADO ABOUT ‘*N*’OTHING[1]

ABSTRACT

This article argues that a core area of the philosophy of biology—the philosophy of fitness—has for decades rested on fundamental conceptual and mathematical errors. These errors have been leveraged to support the position in the philosophy of biology known as statisticalism, which holds that biological fitness does not cause evolution, but is merely a kind of statistical summary of evolutionary outcomes. This is opposed to causalism, which holds that fitness is based on (causally efficacious) probabilistic propensities, a position known as the propensity interpretation of fitness. The error I focus on is the idea that fitness depends on population size, n, and because population size is not a causal quantity, fitness cannot be causal. In this paper, I show that fitness is not dependent on n and therefore a central critique of the propensity interpretation of fitness is ill founded.

---

[1] Thanks to Gianmaria Dani, Alejandro Fábregas-Tejeda, Jack Justus, Andra Meneganzin, Samir Okasha, Maya Parson, Elliott Sober, and an anonymous referee for their comments on earlier drafts of this article. Thanks also to the audiences at the Center for Logic and Philosophy of Science (Institute of Philosophy, KU Leuven), the Department of Philosophy at Heinrich Heine University, the Brandonfest (Department of Philosophy, Duke University), the workshop, “The Fitness Concept in Evolutionary Biology: Philosophical and Scientific Aspects” (Department of Philosophy, University of Bristol), and to Evgeny Havkin, Niklas Parwez, and Gerhard Schurz for the helpful discussions. Finally, thanks to Research Foundation Flanders (FWO) for their support (in particular, for grant #G070122N) and KU Leuven Internal Funds (C1), #3H210777.

Here is the story of the philosophy of biological fitness. In the beginning Brandon, alongside Mills and Beatty, created the *propensity interpretation of fitness* (PIF).[2] This account, as the name suggests, takes fitness to be a *propensity* of organisms to survive and reproduce, as opposed to an *outcome*—a tally of offspring—as it was sometimes assumed to be. The PIF did two important things. One is that it offered a way out of the *tautology problem*. The tautology problem is a kind of reductio that goes like this: (1) The theory of evolution by natural selection holds that evolutionary outcomes are caused by variation in fitness. (2) Fitnesses are reproductive outcomes. Conclusion: Since outcomes cannot cause themselves, the theory of evolution by natural selection is tautologous. One could quibble about whether 'tautology' is the right pejorative here, but in any event, the central claim of the theory of evolution by natural selection seems to be eviscerated by one of its foundational concepts.

Hence fitness as a propensity. This rendering appears to save the theory of natural selection by making the PIF a claim about what sort of effects a particular propensity has. Causal explanations are possible, at least as long as one agrees that propensities are causally efficacious. Let us assume they are and see what else the PIF has on offer.

I said that the PIF did two important things. The other was that it provided a framework for quantifying the fitness of organisms.[3] The PIF helps to quantify fitness by dividing the propensity's outcomes into discrete chunks, allowing these chunks to be tallied and fed into mathematical models. The chunks are based on offspring production. The idea is that an

---

[2] Robert N. Brandon, "Adaptation and evolutionary theory*," Studies in History and Philosophy of Science Part A*, 9(3) (September 1978): 181-206; and Susan K. Mills and John Beatty, "The propensity interpretation of fitness." *Philosophy of Science*, 46(2) (June 1979): 263-286.

[3] While organisms are not the only entities that may be bearers of fitness—genes or groups or even species are sometimes considered to bear (and vary in) fitness—I will here focus on organisms and thus avoid wading through the morass of problems in the debates over the "levels" of selection. If you want a peek at these problems, read Markus Eronen and Grant Ramsey, "What are the 'levels' in levels of selection?," *The British Journal for the Philosophy of Science* 72, 2 (June 2025): 495-518.

organism has a probability of having zero offspring, one offspring, two offspring, and so on. These probabilities can then be tallied to quantify the fitness of the organism. The quantification originally offered by Brandon is the *expected number of offspring*. If a wildebeest has a 0.3 probability of having 0 offspring, a 0.2 probability of having 1, a 0.3 probability of having 2, and a 0.2 probability of having 3, then its expected number of offspring—its fitness—will be $(0.3 \times 0) + (0.2 \times 1) + (0.3 \times 2) + (0.2 \times 3) = 1.4$.

This is thus "expected" in the purely mathematical sense. The expectation value of a distribution is its arithmetic mean. Technically speaking, it is the distribution's first moment. It is thus not the outcome we *expect*: 1.4 is not the number of offspring we would expect from any actual organism.

Fitness calculated in this way allows for evolutionary predictions and explanations. If we have a population consisting of two types of organisms—shade-tolerant and intolerant variants of a plant species, say—if the shade-tolerant variant is fitter than the other, and if the types breed true, then we expect the growth rate of the shade-tolerant type to be higher than that of the shade-intolerant type.

So far so good. The PIF makes the theory of evolution by natural selection non-trivial (or non-tautologous) and provides a way of quantifying this propensity, thereby connecting it to population growth rates. It might seem that this would be the end of the story. The PIF solved the problem—time to move on. This is not what happened, however. Instead, a tide of skepticism about the PIF (as originally formulated) swelled, setting the course for decades of work on the philosophy of fitness.

This skepticism has its roots in some mathematical work that biologist John Gillespie[4] published in the 1970s but was discovered by philosophers a decade later. Recall that the PIF was originally quantified by the first moment of the offspring distribution. Gillespie's work seemed to imply that the offspring distribution's first moment does not completely capture fitness. Instead, the second moment—variance—appears to affect fitness. Variance is the spread of the distribution. You can change the variance in a distribution without changing its mean, and

---

[4] E.g., John H. Gillespie, "Natural selection for within-generation variance in offspring number," *Genetics* 76 (March 1974): 601-606.

vice versa. The mean of the offspring distribution surely affects fitness, but what about variance? On the face of it, it seems like variance would be irrelevant to fitness: what matters is offspring production *on average*, not variance in offspring production. This, I will later argue, is exactly right: variance does not matter. Nevertheless, philosophers became convinced that variance affects fitness and started to chisholm the original fitness definition in order to accommodate variance effects.

Gillespie's work also implied something more radical. Under some circumstances, the fitness of an organism can be changed by changing *n*, the size of the population in which it resides. This appears causally mysterious. It is like the fragility of a vase depending on how many vases there are (in your city or in the world). This worry about *n* triggered a fusillade of work and helped support an entirely new camp in the philosophy of biology, "statisticalism,"[5] which argues that fitness (as well as selection and drift) are not causally efficacious.[6]

This worry about *n*, I will argue here, is much ado about nothing. Fitness is not modulated by population size, nor affected by variance. This is a radical argument. If sound, it implies that much of the work on the philosophy of fitness over more than four decades rests on a faulty foundation and is rife with conceptual and mathematical errors. As we will see, the belief

---

[5] For foundational papers by the statisticalists, see: Mohan Matthen and André Ariew, "Two ways of thinking about fitness and natural selection." *The Journal of Philosophy* 99, 2 (February 2002): 55-83; Denis M. Walsh, Tim Lewens, and André Ariew, "The trials of life: Natural selection and random drift," *Philosophy of Science* 69, 3 (September 2002): 452-473.

[6] While dependence on *n* has helped support statisticalism, it is not the only reason offered in its support. See Denis M. Walsh, André Ariew, and Mohan Matthen, "Four pillars of statisticalism," *Philosophy, Theory, and Practice in Biology* 9, 1 (2017): 1-18. For an early argument for the epiphenomenal character of fitness, see Denis M. Walsh, "Chasing shadows: natural selection and adaptation," *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 31, 1 (2000): 135-153. For a critique of fitness epiphenomenalism, see Lawrence A. Shapiro and Elliott Sober, "Epiphenomenalism-the do's and the don'ts'," in Peter K. Machamer and Gereon Wolters, eds., *Thinking about Causes: From Greek Philosophy to Modern Physics* (Pittsburgh: University of Pittsburgh Press, 2006), pp. 235-264.

that *n* affects fitness is based on the idea that variance affects fitness. Let us thus begin by considering variance before moving on to *n*.

I. VESTIGES OF VARIANCES

In the late 1980s, philosophers, such as Beatty and Finsen[7] and Brandon[8], became convinced that variance in the offspring distribution has a negative effect on fitness: hold the arithmetic mean of the distribution constant, increase the distribution's variance, and the result will be a decrease in fitness. In light of this, Brandon modified his 1978 account of fitness as expected number of offspring to expected number of offspring *minus some function of variance*. He did not specify what that function was but nevertheless felt certain that variance has a negative effect on fitness.

While the source of the idea that variance depresses fitness came from Gillespie's articles, philosophers have typically not engaged with Gillespie's math and instead used toy examples to purportedly illustrate that growth rates decrease with increased variance. Here is a frequently offered example: Take a population composed to two types, *A* and *B*. Every *A* individual invariantly produces two offspring each generation. Thus, each subsequent generation is calculated by multiplying the number of *A*s by 2. By contrast, *B* individuals all produce 1 offspring, or all produce 3 offspring (with equal probability). Each subsequent *B* generation is thus calculated by multiplying the number of *B*s by either 1 or 3. Each *A* has the same expected number of offspring as each *B* (since the arithmetic mean of 1 and 3 is 2), but do they have the same fitness? Here are ten generations of the *A* type (starting with a population size of 1):

$$2 \rightarrow 4 \rightarrow 8 \rightarrow 16 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 256 \rightarrow 512 \rightarrow 1024$$

Since *B* is stochastic, a legitimate comparison would include all possible sequences of generations (producing 1 or 3 offspring) and then average over these possibilities. In the

---

[7] John Beatty and Susan Finsen, "Rethinking the Propensity Interpretation: A Peek Inside Pandora's Box," in: Michael Ruse, ed., *What the Philosophy of Biology Is*. Nijhoff International Philosophy Series, vol 32. (Springer, Dordrecht, 1989).

[8] Robert N. Brandon, *Adaptation and Environment* (Princeton University Press, 1990).

philosophical literature, however, single trajectories are often used.[9] Here is one such trajectory, which starts with 3 then deterministically alternates between 1 and 3:

$3 \to 3 \to 9 \to 9 \to 27 \to 27 \to 81 \to 81 \to 243 \to 243$

After ten generations, there is more than four times the number of *A* individuals as *B* individuals. This trajectory is not a fluke: on average the growth rate of the *B* type will be much lower than that of the *A* type. It seems to follow from this that variance decreases fitness, that Brandon and others were right to see variance as a fitness-depressing factor.

What is clear in this case is that *A*s are fitter than *B*s. What is less certain is whether this fitness difference is *because of variance*. To justify the claim that variance is responsible for the fitness difference, one must compare low and high variance types that are *fair* comparisons. Consider a clearly unfair comparison. If the low variance type always produces 2 offspring, while the high variance type produces 0 or 1 (with equal probability), the high variance type will have lower fitness, but its lower fitness can be attributed to its low average reproductive rate. Even if the low variance type gets lucky and always produces 1 offspring, it will necessarily have a lower growth rate than the low variance type. We obviously judge this to be an unfair comparison. To make this judgment, we are using some kind of criterion of fairness. But what is this criterion?

Is always 2 vs. 1 or 3 (with equal probability) a fair comparison? It *seems* fair because the arithmetic mean of 1 and 3 is 2. Thus, one proposal for a criterion of fairness is to control for the arithmetic mean. But this only leads us to question whether the arithmetic mean is the appropriate criterion. If fairness comes from controlling for a mean, why is the *arithmetic* mean the correct one to control for in this case? There are other means. In particular, there is the geometric mean. The arithmetic mean is an *additive* mean: you add n values then divide by n. The geometric mean is a *multiplicative* mean: you multiply n values then take the n^{th} root. Which mean should we control for?

---

[9] This undermines the comparison, but I will keep to the norms of the literature. I have made sure that the trajectory used here is close to the average, allowing for the desired inferences in this case.

The answer is: it depends. It depends on which kind of model is used for population growth (and, as we will see, different kinds of models are appropriate for different kinds of variance). In the examples above, population growth is modeled as a *multiplicative* process: the population size for the subsequent generation is calculated by multiplying the current population size by a number. This contrasts with *additive* processes, for which subsequent generation population sizes are calculated by *adding* the current population size to a number. For example, we might always add 2 individuals to the population. Or we could flip a coin and add 1 or 3. For a comparison to be fair, I will now argue, one must control for the mean that is appropriate to the model at hand.

In the above trajectories, we created the high variance comparison by controlling for the *additive* (arithmetic) mean, but calculated population growth by *multiplying* the population by a number. Here is what happens when we do this. Start with a zero-variance type—always multiply by x—and create a high variance type by adding and subtracting a value, y, from x. This gives us a comparison between always x vs. $(x - y)$ or $(x + y)$ (with equal probability). Using a multiplicative model to calculate growth rates, the increase in the low variance type across two generations can be calculated by $x \times x$, which is x^2. The high variance type can (on average) be calculated by $(x + y) \times (x - y)$, which is $x^2 - y^2$. Because y^2 is always positive, the high variance type automatically has a lower growth rate if we control for the additive mean with a multiplicative process. Because of this, 1 or 3 (with equal probability) is *not* a fair comparison with always 2 in the case of a multiplicative model.

By comparison, see what happens when we control for additive means for additive processes and multiplicative means for multiplicative processes. For the additive case, the low variance type involves always adding x individuals to the population, whereas for the high variance comparison, either $(x + y)$ or $(x - y)$ (with equal probability) is added. Over two generations of the low variance type, we have $x + x$, which is $2x$. If we sum $(x + y)$ and $(x - y)$, we also get $2x$. Here are sample trajectories (again, starting with a population of 1) where x is 2 and y is 1. The low variance (always add 2) case is:

$3 \rightarrow 5 \rightarrow 7 \rightarrow 9 \rightarrow 11 \rightarrow 13 \rightarrow 15 \rightarrow 17 \rightarrow 19 \rightarrow 21$

And a high variance case (add 1 or 3 with equal probability) sample trajectory is:

4 → 5 → 8 → 9 → 12 → 13 → 16 → 17 → 20 → 21

The high and low variance types have the same growth rates, implying that growth rates are not affected by variance.[10]

For the multiplicative process in which the low variance type is *always multiply by 2*, a high variance fair comparison would be *multiply by 1 or 4 with equal probability* (since the geometric mean of 1 and 4 is 2). The low variance trajectory is the same as the one above, which resulted in a population size of 1042. The high variance type would go like this:

4 → 4 → 16 → 16 → 64 → 64 → 256 → 256 → 1024→ 1024

The high and low variance trajectories have the same outcome. Variances leave no vestiges in growth rates.

II. GILLISPIE'S VITAL DISTINCTION

Because the idea that fitness values can be changed by changing variances or *n* traces back to Gillespie's work, we should pause to consider how he achieved his conclusions. To understand Gillespie, we need to start by introducing a distinction important to his work, that between *within-generation variance* and *between-generation variance*. In the multiplicative examples above, all the organisms of the same type within each generation had the same number of offspring. If one *B* had 3 offspring, every *B* in that generation did so. All the variance is *between* generations, not *within* generations. This synchronized variance, also known as *aggregate variance*, contrasts with cases in which each organism effectively flips a coin and has 1 or 3 offspring. Such unsynchronized, *within-generation variance* is also known as *idiosyncratic variance*.

These two kinds of variance have important implications for how population growth can be calculated. With between-generation variance, we can calculate the subsequent generation by simply *multiplying* the population size by a single number. In the low-variance example above,

---

[10] Again, this is a single trajectory of a stochastic process and thus a comprehensive comparison would need to take into account all possible trajectories, not merely a sample trajectory. Like in the above example, I will not bother with this complication—but rest assured that the sample trajectory has the same qualitative outcome as the average over all trajectories.

we multiplied the population size by 2 to generate the size of the next generation. But for within-generation variance, we cannot do this. If each organism flips a coin to decide if it will produce 1 or 3 offspring, we cannot multiply the population size by a number to generate the subsequent generation's population size. Instead, we need to *add* individuals to the population—in this case adding 1 or 3 for each *B* individual.

What Gillespie argues is that for cases of between-generation variance, fitness is modeled by a multiplicative mean: the geometric mean.[11] For between-generation variance, then, we can use multiplicative models—and to create a fair comparison to see how variance affects population growth, we should control for the geometric mean. The 2 vs. 1 or 4 (with equal probability) case above exemplifies the fact that between-generation variance in offspring number does not affect fitness. Similarly, for within-generation variance, we should use an additive model (and thus the arithmetic mean) to make fair comparisons. As we saw above, within-generation variance has no effect on growth rates (as demonstrated by controlling for the arithmetic mean).

This is not the end of the story, however. For the case of within-generation variance, Gillespie argued that fitness is modeled by the arithmetic mean *minus variance divided by n*. How can population size possibly have anything to do with growth rates? The answer is *it does not*. But to see this, we need to understand what biologists sometimes term *relative fitness* and how it is connected to what I am calling *enrichment and dilution*.

III. ENRICHMENT AND DILUTION

Put a drop of ink into a glass of water. Put a second drop in and you doubled the amount of ink. Put a third drop in and you have increased the amount by only 1.5×. Continue adding ink. The 100th drop increases the amount of ink by a mere 1%, a quantity too small to notice. The effect of

---

[11] The equation he uses to represent the geometric mean—the arithmetic mean minus variance divided by two—is in fact an approximation of the geometric mean. This is an approximation developed before the time of computers in order to predict the performance of stocks in the stock market using arithmetic means and variance alone. (See, William H. Jean and Billy P. Helms, "Geometric mean approximations," *Journal of Financial and Quantitative Analysis* 18:3 (September 1983): 287-293.) It is important to realize that the geometric mean is not a function of variance, as this approximation seems to imply.

the ink thus changes depending on how much has already been added. This creates an asymmetry: adding a drop has a smaller effect than taking one away (if this were possible).

The same phenomenon is true of collections of items. If you have a jar of coins, which contains mostly pennies but a few dimes, adding a dime will make a larger difference to dime frequency than adding a penny will make to penny frequency. And the loss of a dime has a larger effect on dime frequency than the addition of one. If you were to put one dime and one penny into the jar each day, the frequency of the dimes would increase rapidly at first but then eventually converge toward the frequency of the pennies. Analogizing this to biological organisms, one should not interpret the dimes as having a fitness advantage (an advantage of being rare). To do this is to conflate the growth process (a new dime each day) and the frequency changes due to dilution and enrichment. Frequency changes are a function of both causal components (tied to population growth) and noncausal (dilution and enrichment) effects.

Now consider a population of organisms composed of *A*s and *B*s. In the case of within-generation variance, the asymmetry of dilution and enrichment—dilution having a larger effect than enrichment—can give the illusion that variance in offspring number is disadvantageous. When an organism has an offspring, it enriches the population with its type, just like with the coin example. If *B*s produces 1 or 3 with equal probability, if we compare this with always producing 2 offspring, then it seems like always producing 2 carries a fitness advantage. The reason is that having one more than 2 (i.e., having a third) will produce less of a frequency effect than the loss entailed from having 1 fewer. (Again, dilution is more significant for frequency changes than is enrichment.) Thus, *A*s appear fitter than *B*s. But to conclude that *A*s are fitter is to conflate fitness (and the associated growth rates) with dilution and enrichment (and the associated frequency changes).

To bring this distinction into focus, consider fragility. The glass tumbler you drank from this morning has some degree of fragility. If it breaks, it turns into shards. We could quantify the fragility of the tumbler in terms of its expected number of shards after some time (a decade, say). The expected number of shards for the tumbler is due to an array of causally relevant factors, including its intrinsic features (such as its shape and thickness) and features of its environment (how and how frequently it is used).

Now consider the impact the tumbler will have on the global frequency of shards. If the tumbler is clear, it will have an insignificant impact on global clear glass shard frequency. But if the glass is an extremely rare shade of chartreuse called c21, the glass could have a significant impact on c21 shard frequency. The impact the glass has on global c21 shard frequency is not a causal impact. It is simply a numerical impact, one based on three numbers: the global number shards (of any color), the global number of c21 shards, and the number of shards from your tumbler. The causal underpinning of the expected shard frequency of the tumbler is in no way called into question by the fact that the impact it has on shard frequency is due to the number of shards the world over. The mistake of thinking it is, as we will see, is precisely the mistake made by many philosophers of biology.

IV. MUCH ADO ABOUT ‘N’OTHING

The position in the philosophy of biology known as *statisticalism* became established more than two decades ago. Statisticalists reject the idea that the core concepts in evolutionary theory—fitness, selection, drift—pick out causal properties or processes. For statisticalists, it is not true that albino forms of animals are rare because albinism is selected against (where ‘because’ is understood causally). And elephants are not large because of the fitness advantages conferred by their massive size. Instead, fitness and selection denote mere statistical descriptions of evolutionary outcomes. Statisticalists are distinguished from causalists, who tend to view fitness as a probabilistic propensity to survive and reproduce.

An important line of support for statisticalism is based on Gillespie’s equation for fitness in the case of within-generation variance (fitness as the arithmetic mean of the offspring distribution minus variance divided by n). Statisticalists argue that fitness is not causal since “facts such as population size—which is not a causal property of an individual at all—may influence whether a trait’s *frequency* will increase or decrease. […] For a large fraction of organisms, the explanation for *why a trait will increase or decrease* depends in part on factors that are extrinsic to the causal properties of [the] individual.”[12] How are these populational changes related to individual-level propensities? They hold that “these evolutionary dynamics

---

[12] André Ariew, and Zachary Ernst, “What fitness can’t be,” *Erkenntnis* 71, 3 (July 2009): 289-301, here page 296, emphasis added.

are, in fact, counterexamples to the propensity interpretation of fitness."[13] Thus, they use features of (populational) trait *frequency* changes to argue against underlying (organismic) *dispositions*.

Using populational frequency dynamics to argue against the PIF is like claiming that shard frequency's dependence on *n*—global shard number—serves as a counterexample to the fragility propensity. Frequency changes necessarily have noncausal components since they depend on nonlocal factors. Glass shards in rural Indiana, Parisian nightclubs, and Mumbai dumps all bear on global shard frequencies. But they do not serve as counterexamples to the fragility of your tumbler. Part of the diagnosis for why statisticalists appear to think they do is the fact that they do not always clearly distinguish noncausal frequency effects from the causal effects of individual propensities. The first quote in the previous paragraph starts by discussing trait *frequency* changes, but then claims that "why a trait will increase or decrease depends in part on factors that are extrinsic to the causal properties of [the] individual." This quote is ambiguous. If it refers to increases or decreases in frequencies, it is trivially true. If it is about the *number* of individuals with the trait, it is false. The changes in the absolute numbers of pennies and dimes stay the same (one coin added per day) even as relative frequencies change with each additional coin. The fact that frequencies depend on total coin number does not imply that the growth in the number of coins has the same dependence.

In other examples, the statisticalists clearly make the unwarranted leap from frequency to number. For instance, they note that, "demographic factors, such as population size […] also affect the rate of change in *frequency* of a trait type."[14] Two paragraphs later, they conclude that "these considerations and others […] suggest that the *rate of growth* of a trait type is dependent upon all manner of causal, demographic, and statistical factors that are extraneous to the average reproductive output of the individuals possessing a given trait." While the first quote is an observation about trait *frequencies*, the second is about the *growth* of number of individuals bearing the trait. The truth that frequencies depend on global parameters such as *n* in no way implies that changes in the number of individuals with the trait is similarly affected. This

---

[13] *Ibid*, p. 289.

[14] Walsh, Ariew, and Matthen, "Four pillars of statisticalism," *op. cit*., here page 10, emphasis added.

conflation of frequency changes and populational growth allows them to take truths about frequencies and apply them to growth rates, converting them to falsehoods.

To see the absurdity of this conflation, think back to the coin example. The growth of the pennies and dimes is one per day. Because of dilution and enrichment, the equality of growth in absolute coin number will not be reflected by an equality of frequency changes. The rare type will increase in frequency and the dominant type will decrease in frequency—all in the face of an equivalent number of added coins.

V. EXCUSES

Why is there so much confusion about fitness? Why do philosophers fail to clearly distinguish frequency changes from absolute number changes? These are fundamentally different—and only the former depends on *n*. Part of the problem stems from inconsistencies in biological treatments of fitness. Biologists often label fitness as 'absolute fitness' and the metric linked frequency changes as 'relative fitness'.[15] This makes it seem as though there are simply two flavors of fitness, absolute and relative. This is highly misleading. Absolute fitness is causal. It concerns growth rates and that which causes them. Relative fitness is a combination of (causal) absolute fitness and (noncausal) populational components. Even more confusing, sometimes biologists label absolute fitness as *reproductive success* and relative fitness as *fitness*.[16]

The fact that Gillespie used the term 'fitness' and the label 'F' made it seem like he was talking about the same quantity that the PIF theorists were trying to explicate. But Gillespie's equations were concerned with relative, not absolute fitness. Biologists also talk of 'fitness in

---

[15] 'Relative fitness' is variously defined in the literature, depending on what absolute fitness is measured relative to. For instance, it could be relative to the absolute fitness of the fittest individual or to the average absolute fitness. Note that with the latter, relative fitness is based on the fitness values of all the individuals in the population and is thus sensitive to frequency changes. It is this way of characterizing relative fitness that is assumed here. However, no matter how relative fitness is defined, it is important to see that it is a statistic based on absolute fitness values.

[16] For example, Steven A. Frank "Natural selection. I. Variable environments and uncertain returns on investment," *Journal of Evolutionary Biology* 24, 11 (November 2011): 2299-2309.

finite populations' as though fitness has a different character depending on the population size. Again, this is about relative fitness, not absolute fitness.

Had it been clear to philosophers that Gillespie was concerned with relative fitness, not the absolute fitness of the PIF, perhaps the history of the philosophy of biology would have taken a very different course. Consider again his equation for fitness in the case of within-generation variance. Gillespie holds that fitness is the arithmetic mean of the offspring distribution minus variance over *n*. This equation clearly contains the causal component (the expected number of offspring, exactly as described by the original rendering of the PIF) and a noncausal component (the negative effect of the dilution-enrichment asymmetry). This negative factor is an attempt to translate absolute fitnesses into relative fitnesses. The existence of this factor does not serve as a counterexample to the causal fitness component any more than expected shard frequency serves as a counterexample to fragility.

Some biologists are clearer about the relationship between the PIF and the noncausal factors needed to derive expected frequency changes. Hansen,[17] for instance, analyzes the within-generation fitness equation of Gillespie and concludes that the left component (the expectation value) is absolute fitness. The whole factor—including the variance over *n* component—aims to capture relative fitness. However, Hansen shows that the equation is mistaken, that it does not include the variances of all the individuals in the population. This is a fixable problem, however. The details of the fix need not concern us here, but what is clear is that one can subtract the noncausal dilution-enrichment asymmetries from absolute fitness to deduce relative fitness. Again, this does not undercut the causal nature of fitness. All it shows is that relative fitness takes into account populational (or global) factors.

Thus, there are good excuses for confusing changes to global quantities (like relative fitness or frequencies) and absolute number changes. With this diagnosis, it is easy to see that the statisticalist presupposition that Gillespie's within-generation equation serves as a counterexample to the PIF rests on fundamental errors. Nor does Gillespie's work mean that fitness is somehow "holistic"—that it depends on the properties of the whole population, even

---

[17] Thomas F. Hansen "On the definition and measurement of fitness in finite populations," *Journal of Theoretical Biology* 419 (April 2017): 36-43.

ones with no causal influence, as Sober famously suggested.[18] One needs to carefully distinguish the causal from the noncausal properties and acknowledge that the causal, fitness-based component is not called into question by the vagaries of population size or frequency changes.

VII. CAUSES, REAL AND PHANTOM

Some species are polymorphic, meaning that they possess distinct types, such as wing pattern types in a butterfly species. This fact was assumed in the toy examples above, which considered the fitnesses of types *A* and *B* belonging to the same species. In these examples, there was a fixed fitness consequence to being an *A* or *B* and it was argued that this fitness consequence was independent of the frequencies of the types. The nature of dilution and enrichment makes it seem that rare types gain a fitness advantage merely by being rare, but this is an illusion. The advantage of being rare is a phantom cause, not a real one. It is due merely to the numerical asymmetry of dilution and enrichment.

By pointing out this illusion, it might seem that I am arguing against the possibility (or even coherency) of *frequency-dependent selection*, in which the fitness values of the types depend on their frequency. This is not the case, however. I admit the possibility—and even the importance—of frequency-dependent selection in nature. How does this not contradict the argument above that expected frequencies and fitness should not be conflated? Being rare gives you a *frequency advantage* in the sense that an offspring of a rare type makes a larger impact on frequencies than does an offspring of a common type. But this is not an absolute fitness advantage: it has no impact on growth rates and is not causal. Frequency-dependent selection, by contrast, involves a *causal* dependency on frequencies.

One phenomenon that frequency-dependent selection is invoked to explain is the maintenance over time of polymorphisms. Stable polymorphisms are evolutionary puzzles. If the types vary in fitness, then the polymorphism should be short-lived, since the higher fitness type will displace the others. If they are selectively neutral, we expect random drift to eventually lead to one type alone remaining. There are, however, many stable polymorphic species and in some

---

[18] Elliott Sober "The Two Faces of Fitness," in Diane B. Paul, Costas B. Krimbas, John Beatty, and Rama S. Singh, eds., *Thinking about Evolution: Historical, Philosophical, and Political Perspectives* (Cambridge: Cambridge University Press 2001), pp. 309-21.

cases scientists have offered *apostatic selection* as an explanation of their maintenance. Apostatic selection is negative frequency-dependent selection, meaning that there is a selective advantage to being rare. How is apostatic selection causal and not merely an artifact of the dilution-enrichment asymmetry?

Rareness seems to carry an advantage since if we focus on frequency changes, an offspring of a rare type will have a larger frequency impact than will the offspring of a common type. However, what is important for maintaining polymorphisms is the *growth rate* of each type. Growth rates are about the growth of each type, not their relative frequencies, and they do not depend on n. How, then, is it that growth rates can be partially determined by apostatic selection? Let us consider a concrete case to see this clearly.

Many species of salamander exhibit polymorphisms. Scientists have proposed that apostatic selection explains the persistence of the multiplicity of types in these species. Birds are a major predator for salamanders. When birds search for prey, they do so based on a *search image*. The search image for a type is like a concept built from experience with the type. What scientists have found is that birds more easily find common salamander types, and that this is because they have had more exposure to individuals of these types and thus have developed a more robust search image.[19] The causal flow thus goes like this:

exposure to salamanders → search image → predation

Frequencies are causally important since they determine the relative robustness of the search images. For each bird, the total number of salamander exposures of each type helps determine the robustness of the search image of that type. More exposure will lead to a more robust search image, which will allow the birds to more easily find the salamanders, which leads to them prey on the more common type at a disproportionally higher rate.

This example makes clear what frequency-dependent selection is and how this form of selection in no way undermines my argument that we must be careful to distinguish fitness from expected frequency changes. More importantly, it shows that to understand frequency-dependent

---

[19] Benjamin M Fitzpatrick, Kim Shook, and Reuben Izally "Frequency-dependent selection by wild birds promotes polymorphism in model salamanders." *BMC Ecol*ogy 9, 12 (May 2009).

selection, we need to see that rareness alone does not confer a selective advantage. Simply being rare does not trigger apostatic selection. Apostatic selection occurs only when being rare triggers a causal chain that disproportionately selects against the common form.

VIII. THE TYPE TRAP

Many species, such as the salamanders just discussed, are polymorphic. Polymorphisms occur when there are discrete trait types in the population. When traits come in discrete types, they are labeled *qualitative* traits. Such traits are distinguished from *quantitative* traits, which are continuously variable. It is not that Darwin had a height, it is that his height was 183cm. Variables like height and weight are common in nature and are selectively important. In some cases there can be a linear selective gradient (taller is better), while at other times the Goldilocks scenario is realized (average height is best). A selective gradient can prompt directional evolutionary change, whereas with the Goldilocks scenario, selection is said to be *stabilizing* and can help *prevent* evolution from occurring. In either case, this is selection and an evolutionary response—all in the absence of types. And without types, we do not have frequencies.

This implies that one should avoid falling into the trap of thinking that all evolution concerns selection on types, and that fitness is cashed out only in terms of type frequency changes.[20] The PIF posits that organisms have probabilistic propensities of having offspring, and that this serves as the basis for natural selection. Heritable variation in fitness values can cause evolutionary responses—there is no need to resort to types or their frequencies in modeling, explaining, and understanding evolution.

This is not to deny that types play useful roles. They can, especially when the focus is on a phenotypic difference due to a single genetic difference. In the absence of such monogenetic traits, types can also be useful. In some cases, we can sort tokens into types to simplify evolutionary modeling or to more easily generalize. For instance, we could create various types of height. Every 10 cm span (starting at 0 cm) could be a type. Darwin would be height type 19

---

[20] For a recent defense of the idea that types, not tokens, are foundational to natural selection, see Marshall Abrams, *Evolution and the Machinery of Chance: Philosophy, Probability, and Scientific Practice in Biology* (Chicago: University of Chicago Press, 2023).

in this case. But to theorize the selective advantage of belonging to one height type instead of another (19 vs. 18, say) is merely to coarse grain a continuous variable.

Similarly, glasses vary continuously in their thickness. In general, thicker glasses will carry a lower expected number of shards. Like with height, we could create types of glasses. Ones 0-1 mm thick, 1-2 mm thick, and so on. We can then track the frequency of these types. We might gain some degree of predictive accuracy by knowing what type a glass falls into, but there is no certain link between thickness and fragility. Other factors, such as shape and kind of glass—borosilicate, soda-lime, lead crystal—bear on fragility. It is only the full package of specific traits that determines fragility.

The same is true with organisms. Being aggressive without the strength to back it up may be a liability, whereas with the strength, it may be an advantage. This again points to fitness being a property of token organisms, one based on their suite of traits and the environment they find themselves in. While sorting organisms into types can be useful in some circumstances, especially in polymorphic species, it needs to be understood that this is merely for metrological and modeling convenience.

The secondary nature of types further undermines the statisticalist's argument that because the expected frequencies of types depend on *n*, fitness cannot be understood to be a probabilistic propensity of organisms to survive and reproduce. Sorting organisms into types has many uses—but leveraging features of type frequency changes to attempt to undermine the causal effectiveness of the underlying tokens is not one of them.

IX. CONCLUSIONS

Philosophers of biology have for decades been convinced that the fitness of organisms is not fully captured by their expected number of offspring. Some have been persuaded that variance in the offspring distribution depresses fitness and have offered modified mathematical formulae to accommodate offspring variances. And some have been persuaded that *n*, population size, affects fitness, and have proposed a new metaphysics for evolution by natural selection, one that takes fitness and selection to be causally ineffective: 'fitness' is the name for a kind of statistical summary of evolutionary outcomes, not an identifier of that which causes evolution.

I have argued that neither variance in the offspring distribution nor *n* has any effect on fitness. They do impact frequencies, but this is not surprising. Frequencies automatically include *n* (since frequencies necessarily concern the total number of items) and the dilution-enrichment asymmetry links variances with frequency changes. But frequencies are not fitnesses. Realizing this leads to the conclusion that what is true of frequencies is not necessarily true of fitnesses. Total population size is not a fitness component. The same with fragility understood as expected number of shards. Global shard frequencies will be changed upon breaking a glass, but the fragility of the glass is not determined in any way by its impact on global shard frequency.

This article thus shows the fragility of a key pillar of the statisticalists's position. The causalists, however, do not go unscathed. Their efforts to chisolm fitness definitions to fit Gillespie's math fracture upon carefully distinguishing relative frequencies from the absolute fitnesses. My purposes here are not merely destructive, however. My hope is that these insights can be used not just to smash old arguments, but to forge a new understanding of fitness and selection and their place in the theory of evolution by natural selection.

GRANT RAMSEY

KU Leuven