The Conceptual Foundation of the Propensity Interpretation of Fitness

Zachary J. Mayne[1]

[1]Department of History and Philosophy of Science, University of Pittsburgh, 1101 Cathedral of Learning, Pittsburgh, PA, USA.

Contributing authors: zachary.mayne@pitt.edu;

Abstract

The propensity interpretation of fitness (PIF) holds that evolutionary fitness is an objectively probabilistic causal disposition (i.e., a propensity) toward reproductive success. I characterize this as the conceptual foundation of the PIF. Reproductive propensities are meant to explain trends in actual reproductive outcomes. In this paper, I analyze the minimal theoretical and ontological commitments that must accompany the explanatory power afforded by the PIF's foundation. I discuss three senses in which these commitments are less burdensome than has typically been recognized: the PIF's foundation is (i) compatible with a principled pluralism regarding the mathematical relationship between measures of individual and trait reproductive success; (ii) independent of the propensity interpretation of probability; and (iii) independent of microphysical indeterminism. The most substantive ontological commitment of the PIF's foundation is to objective modal structures wherein macrophysical probabilities and causation can be found, but I hedge against metaphysically inflationary readings of this modality.

Keywords: expected fitness, evolutionary explanation, objective probability, causation, modal structure

1 Introduction

The propensity interpretation of fitness (PIF) was independently proposed by Brandon (1978) and Mills and Beatty (1979) to make sense of the ostensible explanatory power that the concept of fitness enjoys in evolutionary theory.[1] On this interpretation, evolutionary entities are understood to have various probabilistic causal dispositions, i.e., propensities, to survive and reproduce (or to *promote* survival and reproduction) in a given environment. This is analogous to the various propensities that assorted household objects have to shatter when dropped. When we resort to the concept of (expected) fitness to explain actual evolutionary trends, we are, in part, explaining what tends to *actually* occur in terms of what is *probable* to occur, much as we do for ordinary games of chance. Additionally, we use what (we suspect) is probable to occur when predicting what will actually occur. We look to natural selection to explain the incredible adaptations that abound in nature because such adaptive traits, thanks to their causal effects, are more likely to propagate than other traits.

The PIF has remained a popular interpretation of fitness through the years, but it has been criticized for having certain theoretical and ontological over-commitments. My aim in this paper is to analyze the minimal theoretical and ontological commitments underwriting the explanatory power that the PIF affords the concept of fitness in evolutionary theory. I will argue that these commitments are less burdensome than has typically been recognized. Toward that end, I will focus my discussion on what I call the *conceptual foundation of the PIF* (CF-PIF). In a slogan, the CF-PIF can be stated as follows: *Fitness is an objectively probabilistic causal disposition toward reproductive success.*

Section 2 explicates why the explanatory machinery of the PIF is constituted by objectively probabilistic causation. The next three sections disentangle the CF-PIF from the most problematic of its alleged over-commitments. Section 3 argues that the CF-PIF is compatible with a principled pluralism regarding the mathematical relationship between measures of individual and trait reproductive success; Section 4 argues that the CF-PIF is independent of the propensity interpretation of probability; and Section 5 argues that the CF-PIF is independent of microphysical indeterminism. Importantly, Section 5 also shows how objective probability and causation can emerge together as *objectively probabilistic causation* from the properties of a modal structure.[2] Section 6 then hedges against metaphysically inflationary interpretations of the objective modal structures to which the CF-PIF is positively committed. Section 7 concludes with a brief summary and reflection on the broader significance of this analysis of the PIF. In addition to its obvious implications for the concept of fitness itself, the following analysis bears on any explanatory domain in which probabilistic causation may serve as an explanans.

[1] Historically, fitness was often identified with actual reproductive outcomes. That is, the organisms or types who actually outcompete their rivals were identified, by definition, as being fitter than those rivals. This is clearly incompatible with fitness's traditional explanatory role, since it reduces to circularity: those organisms or types who achieve the greatest reproductive success do so because they are the ones who achieve the greatest reproductive success (Brandon, 1978).

[2] By "modal structure," I mean that there are facts of the matter as to what would and would not occur in counterfactual scenarios, and that these facts can be given certain structural descriptions. Section 5 provides further elaboration of these ideas.

2 The Explanatory Machinery of the PIF

Though much of the debate over the PIF has centered on fungible details, I am not the first to notice that its core explanatory machinery is quite minimalist. For example, consider Richardson's (1996) review of Brandon's (1990) *Adaptation and Environment.* Richardson notes that, despite Brandon's affinity for the propensity interpretation of probability, the core posits underlying the PIF are really that "expected fitness... is understood as a disposition to manifest some range of reproductive success within a specific environment," and "that the probability in question is an objective probability, independent of actual reproductive success" (Richardson 1996, pp. 125-126; cf. Richardson and Burian 1992). This is almost the CF-PIF. All that I will add to it is an explicit mention of causation, such that the dispositions in question are understood to be causal relationships between physical properties and reproductive outcomes.

The purpose of this section is to explicate these two core explanatory components of the PIF: fitness as *objectively probabilistic* and fitness as a *causal disposition.* They are not independent of one another; indeed, I will argue that the explanatory machinery of the PIF is constituted by *objectively probabilistic causation.* Nonetheless, let us take each component in turn.

2.1 Objectively Probabilistic...

Familiar, though not uncontroversial, points about the explanatory power of objective probabilities in general are relevant here. That my subjective credence in a coin landing heads is equal to 0.51 does not explain why it is that the frequency of heads in a trial approaches 0.51 as the number of coin flips increases toward the infinite limit. Nor does my high credence that the limiting frequency of the coin-flipping setup will be 0.51 explain why this limiting frequency obtains more reliably in longer trials than in shorter ones. Indeed, if there is any explanatory connection, it must run in the opposite direction. If I *should* hold a very high credence that the frequency of heads in longer trials will converge more reliably on the limiting frequency of 0.51 than in shorter trials, then it must be something about the world - about the coin-flipping setup - in conjunction with mathematical principles like the law of large numbers, that constrains my rational credence so.[3]

However, all this is not to say definitively that the stability of limiting frequencies cannot be explained somehow in the absence of objective probabilities. Thoroughgoing subjectivists about probability have their ways of explaining why priors will converge in the limit of gathered evidence (though whether their explanations succeed while staying true to subjectivist principles is another matter). My point is rather that, even if we want to be subjectivists about probability, we should not think that our credences are the thing doing the explanatory work here. To say that fitness is a mere credence is to deprive

[3] These kinds of arguments for objective probabilities run back at least as far as Poincaré (1896), as discussed by von Plato (1983).

it of explanatory power.[4] Those who deny the objectivity of expected fitness therefore ought to find another explanans to replace it.

Perhaps the most prominent view that denies the objectivity of expected fitness is due to Godfrey-Smith (2007, 2009). This view attempts to account for the explanatory power of natural selection while eschewing propensities; it takes recourse only to the concrete causal contributions that trait variation makes to *actual* outcomes, rather than the probabilistic causal relationship between trait variation and *possible* outcomes. Though Godfrey-Smith does not cite a general skepticism of objective probabilities as a motivation for his view, the point remains that fitness cannot do any explanatory work if it is merely a subjective credence. In line with this, Otsuka (2016, p. 478) argues that fitness should retain only a descriptive, rather than explanatory, role in evolutionary biology, while the formal causal structures of evolving populations can inherit the traditional explanatory role of fitness.

The approach to fitness taken by Godfrey-Smith and Otsuka is consistent with my claim that *if* fitness is explanatory, it must be something like a reproductive propensity - they simply reject the antecedent of this conditional by denying that fitness is explanatory. It would be an important task for any full-throated defense of the PIF to motivate the explanatory value of objective probabilities against subjectivist explanatory strategies, and of expected fitness against the explanatory strategy of Godfrey-Smith and Otsuka. Such a task goes beyond the scope of the present paper. My aim is to analyze the theoretical and ontological commitments of the explanatory strategy taken by the CF-PIF, not to decisively knock down all of its competitors. Suffice it to say for now that, since explanatory invocations of fitness obviously do, in fact, pervade the practice of evolutionary biology, we should be hesitant to discard it as an explanans if a plausible physical interpretation can be provided.

Now, return to the analogy between flipping coins and evolving populations. It may initially seem that evolving populations do not obviously manifest stable limiting frequencies in an analogous manner, since each evolutionary "trial" is unique. For most natural populations, we cannot analogously run an evolutionary set-up (viz., a particular population in a particular environment) over and over again in trials of various lengths to see actual outcomes converge on expected outcomes. But we do find an analogy in that, across many independent evolutionary trials, actual outcomes tend to match expected outcomes more reliably in larger populations, as compared to smaller populations. In this sense it is as if, rather than having multiple independent trials of various lengths for a fair coin, we have multiple independent trials of various lengths, each featuring a unique coin with its own unique bias. Just the same as before, the fact to be explained is that the actual frequencies of these trials tend to

[4] See Rosenberg (1983, p. 459) for an opposing perspective that takes the explanatory value of fitness (and probabilities in general) to be of a purely heuristic sort. I resist Rosenberg's characterization of mere heuristics as being genuinely explanatory.

converge toward the probabilistic biases of their respective coins more reliably in longer trials than in shorter trials.[5] A crucial part of the explanatory strategy of invoking propensities is that the biases of the coins are taken to be (estimated) objective probabilities, rather than mere subjective credences alone. So too for fitnesses, which function analogously to the biases of the coins in evolutionary trials.[6]

2.2 ...Causal Dispositions

In addition to there being objective probabilities in the world, the CF-PIF requires that there is causation, such that fitness, taken to be an objectively probabilistic causal disposition, can play its ostensible explanatory role. The explanatory role of fitness is, more precisely, a *causal* explanatory role. The *interventionist* conception of causation (Woodward, 2003) will suffice here. To wit: changes in the physical properties upon which fitness supervenes can make a difference to the probability distributions over evolutionary outcomes.

There is an important subtlety here regarding what exactly serves as the cause and the effect, which can be brought out by attending to a recent discussion from Suárez (2022). Suárez develops and defends a view of fitness as a tripartite 'complex nexus' composed of: (1) observed frequencies or actual manifestations of reproductive success; (2) the probability distributions or expected degrees of reproductive success describing those actual outcomes; and (3) the physical dispositional properties which serve as the supervenience base of the probability distributions. Suárez criticizes the PIF for claiming that dispositional properties *cause* reproductive probabilities, when in fact reproductive probabilities *supervene on* dispositional properties.

This criticism misreads the PIF. The purported causal relationship in the PIF is between material, physical states, not between dispositional properties and probabilities. Namely, what the PIF posits is a causal relationship between physical structures (e.g., organisms, populations, and environments) and reproductive outcomes. It is due to the fact that this causal relationship is an objectively probabilistic one that the term 'propensity' is applicable. Thus, while Suárez is indeed correct that fitness is more complex than just the probabilities that represent it in population-genetics models, this is accommodated well within the bounds of the CF-PIF.[7] There is no reason to suspect, on the basis that fitness is a complex nexus, that fitness is not causal in the sense intended by the PIF. The physical properties upon which fitness differences supervene make a difference to evolutionary outcomes, so fitness is straightforwardly a causal notion in the intended sense.[8]

[5] Larger populations are equivalent to "longer trials" in the sense that, for a type at a given initial frequency, it will take comparatively more sampling instances to push that type to fixation or elimination as we increase the initial population size.

[6] Genetic drift is another crucial component of this story, though just how it fits in is a hotly contested issue. See, inter alia, Clatterbuck et al. (2013); Millstein (2002); Millstein et al. (2009); Sarkar (2011).

[7] Sober (2001) gives much the same analysis of the "two faces of fitness," as Suárez acknowledges in a footnote (Suárez 2022, p. 4).

[8] In interventionist terms: let the relevant physical properties be represented by the variable P, and let reproductive outcomes be represented by the variable R. A counterfactual change in P

To further bracket my construal of causation in the CF-PIF, I will briefly differentiate it here from the superficially similar *causal dispositionalist* version of the PIF, which has been defended by Triviño and Nuño de la Rosa (2022). This view draws on the dispositionalist metaphysics of Mumford and Anjum (2011), which takes dispositions or powers to be irreducible properties of entities that prop up or produce counterfactuals, probabilities, and familiar causal relations. By contrast, I take probabilistic causation to be fully explicable within the empirically accountable modal structures of scientific theories. I fail to see the value in further metaphysical speculation about fundamental entities or properties that might generate those modal structures.

3 Measures of Reproductive Success

The CF-PIF, stated once more: *Fitness is an objectively probabilistic causal disposition toward reproductive success.* The previous section shed light on the two explanatory components of the CF-PIF: objective probabilities and causal dispositions. This section focuses on their explanandum: reproductive success. I will argue that the CF-PIF should be coupled with a principled pluralism regarding the mathematical relationship between measures of individual and trait reproductive success. The "principles" constraining this pluralism are just the usual pragmatic concerns of accountability to scientific practice and coherence under theoretical scrutiny. What we are after is an approach to choosing measures of reproductive success that makes sense of the roles that fitness concepts play in evolutionary biology.

3.1 The Relationship Problem

Since its inception, the measures of reproductive success associated with the PIF have undergone a number of transformations in response to well-known counterexamples. The originally proposed measure of individual fitness - namely, the *expected number of offspring* that an individual will produce in its lifetime - has, in particular, been a prime target of counterexamples. It has been objected that the number of offspring produced by an organism is at times a very poor measure of its reproductive success. For example, delayed sterility can be inherited such that an organism's grandoffspring, great-grandoffspring, or even offspring several generations down the line may predictably fail to reproduce (Pence and Ramsey, 2013; Sober, 2001). We would presumably not want to say that an especially fecund organism is fitter than its conspecifics if it displays such delayed sterility. Similarly, a trait's fitness was originally taken to be the arithmetic mean of the fitnesses (i.e., the expected number of offspring) of the individual organisms who bear that trait, and this mathematical relationship between individual and trait fitness has also been challenged. When there are differences among types (i.e., among groups of individuals who bear a particular trait) in the stochastic variance of offspring-production

would change the probability distribution over the possible values R could take. In a causal graph, we would represent the relationship as $P \to R$ to indicate that P is a causal parent of R.

within a single generation, the associated traits may have different expected frequencies in the next generation even when the expected number of offspring for the individuals who bear each trait are identical (Sober, 2001; Beatty and Finsen, 1989). As population size increases, variance matters less (Ariew and Ernst, 2009; Sober, 2001). Ariew and Ernst take this finding to constitute a fatal blow to the PIF, since it entails that something other than individual reproductive propensities can determine trait fitnesses. That conclusion is far too strong, but the problem is a serious one.

Proponents of the PIF are thus asked: What is the relationship between individual and trait reproductive success, such that the latter may fall out of individual reproductive propensities? I will hereafter refer to this as the "relationship problem." Two prominent solutions to the relationship problem have been developed, so far.

The first solution comes from Pence and Ramsey (2013). They offer a new measure of individual reproductive success, which can be understood roughly as the expected descendant population size as generations go to the infinite limit. The probabilities of trait reproductive success in the infinite long-run can be straightforwardly derived from this measure of individual reproductive success. This indeed solves the relationship problem at face, but it is questionable whether population dynamics regularly conform to the mathematical constraints needed for these measures to be applicable (Suárez, 2022).

The second solution comes from Sober (2013, 2020), who solves the relationship problem by dismissing one of the relata. Sober considerably deflates the importance of individual fitness in evolutionary theory. He notes that measures of individual fitness do not appear in the classic models of population genetics, and that we typically cannot epistemically access individual fitnesses. With individual reproductive success out of the way, Sober then places propensities for the relative reproductive successes of traits directly at the population level. Furthermore, Abrams (2009) argues that type or trait fitness can be given a unitary mathematical characterization. Relative to a pragmatically specified time-interval, type A is fitter than B if A is more likely to increase in frequency than B in the population. Combined with Sober's arguments for the theoretical irrelevance of individual-fitness measures on the one hand and for a population-level propensity interpretation of type-fitness differences on the other, Abrams' argument for the mathematical unification of type-fitness measures would seem to banish the last vestiges of the relationship problem.

However, Pence and Ramsey resist Sober's claim that individual fitness is theoretically irrelevant (Pence and Ramsey 2015; c.f. Pence 2021). They correctly point out that individual fitness plays a fundamental *conceptual* role, even if individual-fitness measures do not appear in the models of population genetics. Whatever it is that traits do to causally promote their own proliferation in a population must ultimately be cashed out in the reproduction of the individuals who bear those traits. References to this conceptual role organismic fitness are commonplace within scientific practice, tracing right back to

Darwin. This is perhaps just another instance of Sober's own "two faces of fitness" (Sober, 2001) unmasking themselves distinctly once more - but then, all the more reason not to eschew individual fitness as theoretically useless. And, given this, it does seem that we should be able to offer up some mathematical measure(s) to fix the reference of such an important concept. The shadow of the relationship problem persists.

3.2 A New Solution: Principled Pluralism

I want to advocate for a principled pluralism regarding the mathematical relationship between individual and trait fitness. I take it that the conceptual role of individual fitness, construed somehow in terms of organismic propensities for survival and reproduction, as (at least part of) the supervenience base of trait fitness is secure for now. I take it that the mathematical unity of trait fitness, as explicated by Abrams (2009), is likewise secure for now. I also take it to be secure that trait fitness (or trait-fitness differences) can be given a population-level propensity interpretation.[9] My principled pluralism manifests itself here as a comfortability with the uncertainty or indeterminacy of the "right" mathematical relationship between individual and trait fitness. There are many degrees of theoretical freedom available in this respect.

Relegating individual fitness to the "conceptual face" of evolutionary fitness alleviates the pressure to produce a mathematical measure of individual fitness from which trait fitness can be straightforwardly derived, e.g., via simply taking an arithmetic mean in all cases. What concrete value, theoretically or philosophically, Pence and Ramsey's infinite individual-fitness measure (or others like it) may offer is still an open question worth pursuing. But even if we retain a simplistic measure of individual fitness, such as the expected number of offspring that an individual will produce in its lifetime, we can still make sense of the conceptual role of individual fitness as an inexhaustive but ineliminable part of the supervenience base of trait fitness. In that case, we would simply have to be prepared to say that the mathematical relationship between trait fitness and individual fitness can sometimes change due to other factors, like population size.

It should not be terribly surprising that the way in which a trait will tend to proliferate throughout an arrangement of individuals sometimes depends not just on the causal properties of the individuals (already understood relative to the population and environment) but also on the structural properties of the way those individuals are arranged, including their number. But even these structural properties become yet more causal properties when we adjust our focus to the population level. Trait fitness (or trait-fitness differences), as a population-level reproductive propensity, thus may still be said to supervene in part on the most simplistic conception of individual reproductive propensities,

[9] The inclusion of the parenthetical here is a nod to Sober (2013), who argues that trait-fitness differences, but not trait fitnesses themselves, are population-level propensities. In the next section, I defend the view that trait fitnesses are population-level propensities.

even after taking stock of all the difficulties facing any consistent and simple mathematical relationship posited to hold between the two.

4 The Propensity Interpretation of Probability

A common objection levied against the PIF is that it inherits fatal flaws from the propensity interpretation of probability (PIP), *sensu* Popper (1959). These objections come in two forms. In the first form, general problems that have been raised for the PIP are cited as reasons to reject the PIF, absent any detailed analysis of how those problems carry over beyond a gesture toward the received view that the PIF relies on the PIP. In the second form, objections that have been inspired by problems known to plague the PIP are repurposed *mutatis mutandis* for the PIF. Bourrat (2017, p.28) raises the first sort of objection. Sober (2013) briefly cites the former sort of objections but (rightly) goes on to emphasize the latter sort. Sober's solution to these problems is to abandon the view that trait fitnesses are propensities, instead holding that only trait-fitness *differences* are (population-level) propensities. I will first argue that the most powerful generic objections to the PIP do not pose any danger to the CF-PIF.[10] Then, I will argue, *pace* Sober, that trait fitnesses should be interpreted as population-level propensities, so long as one conditionalizes on the correct population-level causal properties.

4.1 Generic Objections to the PIP

Consider the argument known as Humphreys' Paradox (Humphreys, 1985, 2004; Salmon, 1979), which has been quite persuasive in showing that propensities cannot be conditional probabilities because they do not obey the Kolmogorov probability calculus. According to the probability calculus, $P(A|B) = P(A)P(B|A) \ / \ P(B)$. Thus, supposing that $Pr(A|B)$ is a propensity for B to cause A, it seems that the probability calculus tells us that there is a well-defined inverse propensity $Pr(B|A) = Pr(B)Pr(A|B) \ / \ Pr(A)$, which is the propensity for A to cause B.[11] But propensities are often not reversible in this way. There is a well-defined propensity for moisture saturation in a cloud at t_1 to cause rain at t_2, but there is no well-defined propensity for rain at t_2 to cause moisture saturation in a cloud at t_1.

Suárez (2013) has noted that this version of Humphreys' Paradox (HP) is only problematic if one holds a particular 'identity thesis': namely, that *all*

[10] Drouet and Merlin (2015) reach a superficially similar conclusion, in that they also claim the PIF does not depend on the PIP. However, they dispense with the explanatory power of causal dispositions and resort to a purely statistical analysis of fitness explanations, which I take to be a rejection of the conceptual foundation of the PIF. The view they defend is not, in this important sense, a variant of the PIF.

[11] Note that, in the literature, both $Pr()$ and $P()$ are commonly used interchangeably to represent probabilities and/or propensities. For clarity in the following discussion of the distinction between probabilities and propensities, I will use the notation $P()$ to represent all conditional probabilities, any of which may or may not represent propensities, and the notation $Pr()$ only for those probabilities that are explicitly claimed to represent propensities.

conditional probabilities are propensities.[12] Following Suárez, I will call this the *Identity*$_1$ thesis. If one takes propensities to be a proper subset of conditional probabilities, then one can perfectly well accept that for some conditional probability $Pr(A|B)$ which is a propensity, the inverse conditional probability $P(B|A)$ is not a propensity. Nothing in the CF-PIF commits us to this identity thesis, so any objections to the PIF that cite such arguments against the PIP without further elaboration are, at best, incomplete.

Suárez (2013, 2022) notes that there is a second identity thesis, *Identity*$_2$, which he claims is undermined by a more sophisticated version of HP involving the quantum-mechanical transmission or absorption of photons through a half-silver mirror (Humphreys, 1985, 2004). This second identity thesis holds that *all propensities are conditional probabilities*. What this sophisticated version of HP alleges to demonstrate directly is that at least some propensities cannot be probabilities, since they do not obey the probability calculus. This is so not because their inverse conditional probabilities seem not to be propensities but because they lead to formal contradictions within the probability calculus.[13]

Suárez's solution to HP, following Humphreys' own preferred solution, is to abandon the *Identity*$_2$ thesis and, along with it, the notion that the Kolmogorov probability calculus is the correct formalism for propensities. Suárez treats propensities as a *sui generis* relationship between a causal setup C and a manifested probability distribution $\mathrm{P}(M_i)$, where M_i are the various possible outcomes or manifestations that could result from the propensity. This *sui generis* relationship is represented as $C \gg \mathrm{P}(M_i)$, and since this is an absolute probability rather than a conditional probability, it seems HP will not apply. As I noted in Section 2.2, it is not clear why the CF-PIF should commit us to the second identity thesis, and thus it seems Suárez's solution to HP is open to the CF-PIF as well.

There is, however, a weaker thesis paralleling the second identity thesis, which threatens to produce some lingering troubles with HP for Suárez and the CF-PIF alike. Call this the *Representation*$_2$ thesis: *all real propensity relationships can be represented as conditional probabilities*, whose values will be equal to the probabilistic manifestation of the propensity when the conditioning events include all and only those events that instantiate the propensity. In other words, for some propensity relationship $C \gg \mathrm{P}(M)$, we can always find a conditional probability $\mathrm{P}(M|C)$ such that $C \gg \mathrm{P}(M) = \mathrm{P}(M|C)$. Suárez seemingly resists *Representation*$_2$ along with *Identity*$_2$ in one fell swoop:

> Propensities are on this view not to be identified with probabilities. Instead they are more generally taken to be dispositional properties

[12] See also Niiniluoto (1988, pp. 103-104), who makes basically the same point in a footnote. Thanks to an anonymous reviewer for bringing this to my attention.

[13] See Humphreys (2004, pp. 668-669) and Suárez (2013, pp. 80-83) for the full formal paradox. Briefly: photons are fired from a laser toward a half-silver mirror at some time t_1; at a later time t_2, the photons hit or miss the mirror with some probability; and at a still later time t_3, the photons are transmitted through or absorbed by the mirror with some probability. The formal contradiction arises within the probability calculus when we combine a number of seemingly reasonable assumptions about the propensity of the experimental setup at t_1 to produce outcomes at later times.

> with probabilistic displays or manifestations. There is on this view no need to represent the relation between the propensity and its manifestations as a conditional probability... (Suárez, 2013, p. 87).

However, to reject *Representation*$_2$ on the basis that we *do not need* to represent the propensity relationship as a conditional probability would be too hasty. For *Representation*$_2$ states only that *there is* some conditional probability that can represent any given propensity relationship, and this may be so independently of whether we choose to use that conditional probability for anything. If one accepts *Representation*$_2$, and if one supposes that all the propensity relationships specified in the sophisticated version of HP are real propensity relationships, then HP runs just as it ordinarily would, and we derive a contradiction.

So, why think that *Representation*$_2$ holds, given Suárez's conception of a *sui generis* propensity relationship? There is initial cause for doubting *Representation*$_2$ when one considers that conditional probabilities are measure-theoretic entities whose values pick out subsets of outcome spaces, whereas propensities, for Suárez and Humphreys, are material events to which the probability calculus does not apply. But suppose that I formulate a conditional probability for some material event M that conditionalizes on all and only those material events C that instantiate the propensity to cause M. I have thereby picked out, in a measure-theoretic way, the subset of the outcome space in which those material events obtain. Within this outcome-space subset, because I know that C obtains, the *sui generis* relationship between C and the manifested probability $\mathrm{P}(M)$ will then force me to assign a particular value to $\mathrm{P}(M)$. But this is just what it is to say that the conditional probability $\mathrm{P}(M|C)$ must take the value of $C \gg \mathrm{P}(M)$, which is precisely the *Representation*$_2$ thesis.

As stated above, the sophisticated version of HP continues to produce a formal contradiction if one accepts both *Representation*$_2$ and the claim that all of the supposed propensities specified by Humphreys represent real propensity relationships. Rejecting both identity theses is insufficient to dispel HP. Since *Representation*$_2$ is correct, let us consider our only other option. There are, I argue, two supposed conditional propensities specified by Humphreys that do not represent any real propensity relationship. Crucial to Humphreys' derivation of a formal contradiction is the following principle of *conditional independence*:

$$\text{(CI) } Pr_{t1}(I_{t2}|T_{t3}B_{t1}) = Pr_{t1}(I_{t2}|\neg T_{t3}B_{t1}) = Pr_{t1}(I_{t2}|B_{t1})$$

In plain English, CI reads, "The propensity (Pr) of the experimental setup at t_1 , together with all of the relevant background information (B), to produce the incidence (I) of the photon on the half-silver mirror at t_2 is unchanged if we conditionalize on the transmission (T) or failure of transmission ($\neg T$) of the photon through the mirror at t_3."

As informally stated, CI is surely correct. Indeed, the *propensity* at t_1 to cause t_2 must be unaffected by anything that occurs after t_2. What we should reject instead is the formalization of CI that appears in the sophisticated version of HP, which inappropriately treats as propensities two conditional probabilities that do not represent real propensities. Since anything that occurs after t_2 cannot be a cause of I, conditionalizing on T or $\neg T$ at t_3 means that our conditional probability no longer conditionalizes on *all and only* those material events that instantiate the propensity, so *Representation*$_2$ no longer tells us that our conditional probability represents (i.e., takes the value of) any propensity. Dropping the inappropriate uses of propensity notation, we can update CI to CI*:

$$(\text{CI*})\ P(I_{t2}|T_{t3}B_{t1}) = P(I_{t2}|\neg T_{t3}B_{t1}) = Pr_{t1}(I_{t2}|B_{t1})$$

Now, keeping in mind that the first two conditional probability terms in CI* do not represent propensity relationships, we can see that CI* is plainly false. The propensity at t_1 to cause incidence at t_2 is between 0 and 1, while the probability of incidence at t_2 given transmission at t_3 (in the measure-theoretic sense of "given") must be equal to 1, since transmission is impossible without incidence. Consequently, all that the sophisticated version of HP definitively undermines is *Identity*$_1$, along with a representation thesis paralleling it, *Representation*$_1$, which states: *all conditional probabilities represent propensity relationships.* But the simpler version of HP already gave us ample reason to reject those two parallel theses. The sophisticated version of HP does not tell us anything new. It just reiterates that not all conditional probabilities are, nor represent, propensities.

The solution to HP with which we are left is surprisingly simple. All we need are ordinary, measure-theoretic conditional probabilities, together with the claim that some (but not all) of those conditional probabilities represent propensity relationships.[14] Specifically, the conditional probability for an event M that conditionalizes on all and only those causal properties C which instantiate the propensity to produce M, i.e., $P(M|C)$, will take the value of the propensity $C \gg P(M)$.

4.2 Specific PIP-Inspired Objections to the PIF

Now, let us consider an argument that is hand-crafted to the PIF. Sober (2013) notes that the expected fitness of a trait (understood as the probability of an organism's survival or reproduction, given that the organism possesses some trait) is often not a measure of the causal power of that trait to produce survival or reproduction. For instance, a causally inert trait may be perfectly genetically linked to a causally efficacious trait in a population. Since the two traits are co-extensive, their trait-fitness values must be identical, and yet the former has no propensity to produce survival and reproduction. Hence, Sober argues that trait fitness is not a propensity. Differences between the fitnesses of two traits, however, are population-level propensities. Even if two traits are causally inert, the linkage of one of those inert traits to a causally efficacious

[14] Section 5 details how objective probability and causation can emerge together from properties of a modal structure.

trait is enough to instantiate the population-level propensity for an increase in frequency of that inert trait.

The general considerations of the PIP explored above will be valuable in understanding why this argument does not, in the end, succeed against the view that trait fitnesses are propensities. The first point worth noting is that, at most, Sober's argument shows that *some* trait fitnesses are not propensities. This is not an insignificant conclusion, but we can go further. I will argue that all trait fitnesses are population-level propensities; but we must conditionalize on the right population-level causal properties to see why. Any change in a trait's expected fitness must be realized by a physical change in the population, whether via individual reproductive propensities, distributions of the trait among individuals, linkage between traits, population size, or whatever else. Just as with trait-fitness differences, it may not be the trait itself which is causal in determining its own degree of success, but said degree will always be causally determined by physical aspects of the population.

Take a causally inert trait X, for which the probability of survival and/or reproductive success (S) of an organism that bears the trait X is equal to $P(S|X)$. Sober is right to note that this conditional probability does not represent the propensity of X to produce S, since X is inert. But now, let L represent the fact that X is co-extensive with a trait Y, which is causally responsible for producing S. We find that if we conditionalize on $X\&L$, our trait expected-fitness value $Pr(S|X\&L)$ does indeed represent a population-level propensity. Note also that $P(S|X) = Pr(S|X\&L)$; the former is merely shorthand for the latter, since L was implicit in our background information all along. What can safely be said, therefore, is that while a trait's expected-fitness value $P(S|X)$ does not necessarily represent the propensity of the trait X to cause the outcome S, it does implicitly represent the propensity of the population to manifest S in those organisms with the trait X. The other material properties of the population that instantiate the propensity have merely been left unnoticed in the background information.

5 Determinism, Macro-Probabilities, and Modal Structure

A commitment to microphysical indeterminism is also sometimes attributed to the PIF (Bourrat, 2017). Of course, the CF-PIF makes no explicit mention of indeterminism. What it does require is treating probability as something that is 'out there in the world,' rather than a mere expression of our ignorance of details, such that it can do real explanatory work on questions regarding why populations evolve in the ways that they do. However, there is a historically commonplace view which holds that objective physical probabilities are incompatible with microphysical determinism. If that view were correct, then the CF-PIF would depend on microphysical indeterminism after all. In this section, I disentangle the CF-PIF from microphysical indeterminism. I do this by connecting a number of different threads in the literature concerning compatibilism between determinism and objective probability. The principal connection between these threads is, I argue, the (explicit or implicit) invocation of objective modal relationships between macroscopic states and the sets of microscopic states upon which the former could possibly supervene. Along the way, I will also show how the interventionist account of causation fits seamlessly into the resulting modal picture, such that objectively probabilistic causation may emerge from the properties of a modal structure.

5.1 Why a Commitment to Microphysical Indeterminism Matters

There is good reason to be hesitant to adopt an interpretation of fitness that commits itself to microphysical indeterminism. The jury is still out on which interpretation of quantum mechanics is correct, and some of the contenders are deterministic (e.g., Bohmian mechanics), even if they represent a minority of views among experts. If the PIF relied on microphysical indeterminism, it would require not only that deterministic interpretations of quantum mechanics are false and that microphysical indeterminism *sometimes* translates to indeterminism at biological scales (e.g., via genetic mutations due to quantum events) but that *all* of the purportedly objective macro-probabilities involved in fitness explanations can be similarly accounted for. Furthermore, even if quantum probabilities are irreducible, there are good reasons to doubt whether they regularly combine in such a way as to produce the particular probabilities that we are interested in at macroscopic scales (Abrams, 2007, p. 15). If the PIF ties itself to this thesis of microphysical indeterminism with ubiquitous macrophysical consequences, therefore, it does so at its own peril. These are the sorts of considerations that motivate me to heed Millstein's call for evolutionary accounts to be compatible with both determinism and indeterminism (Millstein, 2003).

However, arguments for compatibilism between objective physical probability and microphysical determinism have been growing in popularity for quite some time. In fact, compatibilist arguments for deterministic *irreducible chances* are also growing in popularity, though this is a more radical thesis than the one I will defend here.[15] If microphysical determinism and irreducible

[15] It is a radical thesis because irreducible chances are usually understood to rationally constrain one's credence in the occurrence of an event in a principled way. If the irreducible chance of event P occurring is 0.2, one is rationally compelled to set their credence in the occurrence of that event to 0.2, no matter what other information one might learn about the physical world up to and including time t (this is the 'Principal Principle' introduced by Lewis (1980)). The usual argument for incompatibilism regarding determinism and irreducible chance is that, in a deterministic world, the rational credence for an event can always in principle be reduced to 0 or 1 (or made to approach these values as a limit, perhaps) by gathering ever more precise information about the microphysical state of the world. The most vivid illustration of this issue invokes a hypothetical 'Laplacian intelligence,' which can know the microstate of the entire universe with arbitrary precision and calculate the future and past trajectories with perfect computational ability according to the universe's dynamical laws. At face, this seems to entail that irreducible chances must be 'baked in' to the fundamental dynamics of the universe, if any such things as irreducible chances exist. Clever arguments are needed if the in-principle credence-constraining aspects of irreducible chances are to be compatible with determinism. See Ismael (2009, 2011) for a few such clever arguments. See also List and Pivato (2015) for an argument that, if successful, would entail that the

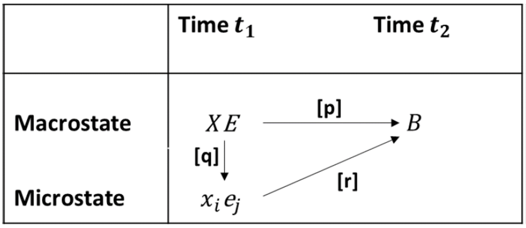

Fig. 1 A reconstruction of Sober's model of macro-probability: [p] = $P(B$ at $t_2|$ XE at t_1); [q] = $P(x_ie_j$ at $t_1|$ XE at $t_1)$; [r] = $P(B$ at $t_2|x_ie_j$ at $t_1)$.

chances turn out to be compatible, then of course determinism and objective probability would be compatible, since irreducible chances are definitionally objective probabilities. But we can have objective probability without irreducible chances, at least in principle, and I think we already have good reason to believe that something other than irreducible chances *does* realize objective probabilities.

5.2 Probability and Causation within Modal Structure

Let us begin with a compatibilist view of determinism and objective macro-probabilities, which has been put forward by Sober (2010). I will introduce some new symbolism here and translate Sober's argument accordingly. Let X be an evolutionary entity, whether an organism/individual or trait/type; let E be the exhaustive causally relevant population and environment within which X is embedded at some time t_1; and let A, B, and C be the three possible reproductive outcomes for X in E that could obtain at a later time t_2. For a given initial macrophysical state XE at t_1, there are many different microphysical states that could realize XE. Let x_ie_j be an arbitrarily chosen microstate from the set of all possible microstates that could realize the macrostate XE. Sober argues that if (i) there is an objective probability [r] of an initial microstate (e.g., x_ie_j) at t_1 causing a resulting macrostate (e.g., B) at t_2 and (ii) there is an objective probability [q] of being in some initial microstate given that our scenario started in some macrostate (e.g., XE), then the macro-probability [p], or $P(B$ at $t_2|$ XE at $t_1)$, is objective (see Figure 1). More precisely, $[p] = \sum_{k=1}^{n}[q]_k[r]_k$, where there are n unique initial microstates. This argument holds even if we assume deterministic relationships between each unique microstate and the resulting macrostate, i.e., if $[r]_k$ can take only the extremal values of 0 or 1.

When it comes to the question of why we should take probabilities like [q] to be objective, Sober advocates for a "*no-theory theory of probability*" which leaves such probabilities as primitive theoretical quantities (Sober, 2010, p. 149). Sober draws an analogy with mass. Although mass is a primitive theoretical quantity, our belief in its existence is justified by the reliable convergence of independent empirical measurements of mass. Similarly, independent

counterfactual macro-probabilities described in this section are, in a meaningful sense, irreducible chances.

empirical measurements of [q], i.e., the probability distribution of possible initial microstates that could realize some initial macrostate, often show reliable convergence. The example Sober gives here is of a coin flip. The frequency of landing 'heads' reliably converges on 51% as the number of coins flipped increases. This empirically supports the existence of an objective probability distribution for [q] and, since [r] is uncontroversially objective even if determinism is true, also therefore empirically supports the objective existence of the macro-probability [p].[16]

What I will now argue is that, rather than terminating our ontological inquiry here in the no-theory theory of probability, we can (and should) conceptualize Sober's macro-probabilities in terms of objective modal structures. Here, I build on a framework from Lyon (2011) called 'counterfactual probability,' which in turn draws on Bigelow (1976).[17] Lyon (2011) describes counterfactual probability as a second kind of objective physical probability, distinct from irreducible chance, which "is a measure of how *robust* a proposition is under a class of counterfactual situations" (p. 429, emphasis in original) and entails "those probabilities in explanations that give some level of modally comparative information" (p. 431).

Explanations that invoke counterfactual probabilities are modally comparative because they concern regularities in how the macroscopic world would tend to look across a range of possible microstates. For instance, counterfactual probabilities are what we invoke in classical statistical mechanics when we state that the microphysical particles comprising macrophysical ice cubes in tepid water tend to dissipate energy in patterns such that the cubes melt at predictable rates. In giving such explanations, we claim that certain systems, in virtue of their macroscopic properties, will tend to presently be in microstates that will evolve into future microstates that realize particular macrostates.

It should be noted that the view I defend here entails only objective probabilities that are optimally *explanatory* of evolutionary trends of a very broad scope, not optimally *predictive* of particular outcomes.[18] Fitness explanations purport to account for why, in virtue of having certain macroscopic traits, some types of organisms tend to outcompete others in their shared macroscopic environment. These explanations are not in the business of stating with maximal

[16] A related point has been raised by von Plato (1983, p. 45) with respect to the *method of arbitrary functions*, in which the production of stable frequencies is explained by the following property exhibited by some spaces of initial microstates: For any continuous distribution on the initial space, and given the dynamics of the system, the same resulting proportion of outcomes will be obtained. As for coin flips, in this respect, so too for evolving populations. For much the same reasons, we have evidence that the structures of the spaces of initial microstates in evolutionary systems, together with the system dynamics, uniquely determine evolutionary outcome probabilities. However, see de Canson (2022) for a discussion of some complications regarding how exactly the method of arbitrary functions relates to system dynamics and objective probabilities.

[17] Bourrat (2017) defends objective deterministic probability by drawing primarily on the 'natural range' conception of probability developed by Rosenthal (2010). See also Abrams (2012) and Strevens (2011) for closely related views. Exploring the differences between Rosenthal's, Abrams', and Strevens' frameworks goes beyond the scope of this paper, but it is quite plausible that all three are classifiable as specific subtypes of the counterfactual probability framework. See also Batterman and Rice (2014) for an approach that is amenable to a "modal robustness" analysis of macrophysical dynamics.

[18] This distinction is inspired by Sober (1984). Macro-probabilities are predictive of particular outcomes to some degree, just not optimally so.

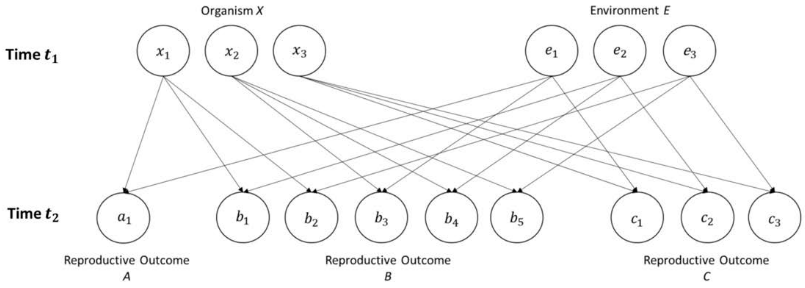

Fig. 2 A diagram of deterministic causal interactions between possible microstates, together with supervenience relationships between possible microstates and macrostates, giving rise to counterfactual probability.

accuracy what this-or-that exact population will actually do in this-or-that exact environment. Of course, attending to the relevant microstates would be optimal in that case, and our use of counterfactual macro-probabilities for such purposes reflects our ignorance thereof. Rather, fitness explanations are in the business of making intelligible some respectably large portion of the modal structure of the biological world. For example, fitness explanations tell us things like "melanism was evolutionarily advantageous among peppered moths in industrial London because, amidst the pollution, it conferred superior camouflage for avoiding predation." This means, among other things, that, given those macroscopic starting conditions, the space of possible scenarios we could find ourselves in is disproportionately constituted by scenarios wherein the melanic moths reproductively outcompete the non-melanic moths in industrial London. In other words, the probability of melanic moths outcompeting non-melanic moths was (objectively) high. Even if we had found ourselves in a scenario where the melanic moths failed to win out, and even if the world were deterministic, that explanation would have remained true.

Lyon does not give a formal analysis of counterfactual probability, but Sober's formalism is apt to play this role. In Figure 2, organism X and environment E are depicted causally interacting to produce some reproductive outcome. Because X and E are macroscopic states, all of their possible initial microstates are listed (x_1, e_1, etc.). This causal model is deterministic, so each unique microstate interaction produces a unique micro-outcome (e.g., the scenario containing initial microstate x_2e_3 at t_1 always also contains micro-outcome b_5 at t_2). Like X and E, the reproductive outcomes (A, B, C) are also macroscopic, so the micro-outcomes can be grouped according to the macro-outcomes that supervene upon them. The counterfactual probability of each possible reproductive macro-outcome is equal to its size, measured as the volume of a subspace of possible micro-outcomes, relative to the volume of the space of all possible micro-outcomes. Macro-outcome A has a probability of $1/9 = 11.1\%$; B has a probability of $5/9 = 55.6\%$; and C has a probability of

$3/9 = 33.3\%$.[19] If we apply Sober's formalism to Figure 2, we get precisely the same result. Indeed, Figure 2 is just a re-imagining of Sober's model in Figure 1 in the form of a causal diagram. The probability [q] is just the probability of being in some causal microstate $x_i e_j$, given the causal macrostate XE; the probability [r] is just the probability of a causal microstate $x_i e_j$ at t_1 causing a particular microstate to obtain at t_2; and the probability [p] is just the macro-probability of the causal macrostate XE causing a particular macrostate to obtain at t_2.

Rather than interpreting probabilities like [q] as primitive theoretical quantities, we should recognize that they emerge from properties of modal structures, in the sense just described. I see two benefits of making this move. First, it clarifies the kind of reasoning already implicit in Sober's analysis, where he looks to the structure of the space of microstates in order to derive a particular value for [q]. Second, it allows macrophysical probabilities and causation to be understood together in a unified framework. I promised at the beginning of this section to show how interventionist causation fits into this modal picture. The work is already done, thanks to our analysis of the causal diagram in Figure 2. If we intervene on X or E to change its macrophysical structure, we will thereby change the space of possible microphysical states upon which XE could supervene, and this, trivially, can change the macro-probabilities of reproductive outcomes. Macrophysical causation thus can supervene on microphysical causation in just the same way that objective macro-probabilities can supervene on deterministic microphysical dynamics;[20] indeed, the two can emerge together from the properties of a modal structure as *objectively probabilistic causation.*

Let us now briefly take stock of what this section has established thus far. First, following Sober and others, we should hold that objective macro-probabilities are compatible with microphysical determinism. Accordingly, the CF-PIF does not depend in any way on microphysical indeterminism. Second, by subsuming Sober's formalism within Lyon's conceptual framework of counterfactual probability, we have seen how objective probabilities and causal relationships can emerge together from properties of a modal structure. As a corollary, note that no metaphysically inflationary posits about fundamental powers needed to be made at the outset. If there are physical facts about what would obtain in various counterfactual scenarios, then we have everything we need for the CF-PIF to do its explanatory work, at least in principle. Namely, we can have objectively probabilistic causation.

5.3 Objections Considered

However, not all conceivable modal structures have the necessary properties for objectively probabilistic causation. For instance, suppose there were infinitely

[19] For simplicity, I assume in this example that the volume of the possibility space of a macrostate splits equally among its possible microstates, but I do not assume that this is always the case in reality. Further discussion of this point is given in Section 5.3.

[20] This falls in line with recent arguments for causal-supervenience theories in biology, such as those given by Boyd (2017) and Pence (2021).

many distinguishable microstates upon which a given macrostate could possibly supervene. In that case, we could not normalize a distribution over the possible microstates while assigning equal non-zero measures for each of their individual modal volumes. Abrams (2006) points out that there could be no well-defined type frequencies in hypothetical infinite populations for much the same reason. A particularly striking illustration of the potential for modal structures to be inhospitable to both probability and causation can be found in the case of Norton's dome (Norton, 2003, 2021). Norton formally specifies a Newtonian world in which a ball sits atop a dome with a very particular geometry. There are two solutions to the dome's dynamical equations. In the first solution, the ball sits motionless atop the dome for eternity. In the second solution, at an arbitrary time t, the ball spontaneously rolls down the dome in an arbitrary direction. The dome-world's possibility space thus includes the ball's eternal motionlessness along with the ball's spontaneous movement in each direction at each time t. This provides no non-arbitrary way to define a probability distribution over the possibilities. Furthermore, since the ball rests or moves completely arbitrarily, no causal story can be told about why it does so.

Sober's analysis of macro-probabilities requires that there is an objective synchronic probability distribution [q] over the space of possible distinguishable microstates, and this, in turn, requires that the space is normalizable. So, the worry may arise that some future fundamental physics will undermine this normalizability and, along with it, the objectivity of macro-probabilities. In response, recall Sober's appeal to the empirical evidence of a limiting frequency that is stable across multiple independent experimental set-ups in the case of flipping coins. If the space of causally relevant microstates were not normalizable, then there would be no objective reason to expect the stable limiting frequency of 0.51 to obtain. In a world filled only with a vast number of Norton's domes, we should not expect to see any such stable patterns of physical behavior. That we do see such stable patterns in coin-flipping, statistical mechanics, and evolutionary biology tells us something empirically: it suggests that the relevant spaces of microstates are normalizable, *at least insofar as the microstates can be distinguished by their ability to make a causal difference to the macroscopic behavior under consideration.*

Another worry that may arise here is more conceptual and less amenable to an empirical response than the last. The worry is that microstates may not always be perfectly sortable into categories like [realizes macrostate A] or [realizes macrostate B], since there may not always be objectively well-defined delineations between macrostates. If macrostate B supervenes on microstate o_k perfectly while macrostate C cannot supervene on o_k at all, but there is no perfectly clear border between B and C, then how is the sharp transition from [r] = 1 for B to [r] = 0 for C to be explained? Presumably, it cannot. Note that what is being expressed here is not the worry that we cannot pick out macrostates in a consistent and objective enough way for them to stand in objective relations to microstates. In response to *that* worry, we could give the

$$\begin{tikzcd}
 & \text{Time } t_1 & \text{Time } t_2 \\
\text{Macrostate} & XE \arrow[r, "{[p]}"] \arrow[d, "{[q]}"'] & B \\
\text{Microstate} & x_i e_j \arrow[r, "{[u]}"] & o_k \arrow[u, "S*"']
\end{tikzcd}$$

Fig. 3 A modified version of the reconstruction of Sober's model of macro-probability shown in Figure 1: [p] = $P(B$ at $t_2|$ XE at t_1); [q] = $P(x_ie_j$ at $t_1|$ XE at $t_1)$; [r] = $P(B$ at $t_2|x_ie_j$ at $t_1)$; S^* = micro-macro supervenience relation.

same empirical response as before. Rather, what is being expressed here is the worry that our formalism may impose sharp delineations between macrostates, which invokes a particular metaphysical view of nature that may outstrip our scientifically and philosophically responsible characterizations thereof.

A potential conceptual/formal solution to this worry is displayed in Figure 3, where I show that [r] can be decomposed into [u] = $P(o_k$ at $t_2|x_ie_j$ at $t_1)$ multiplied by a supervenience relation, S$^*(o_k,B)$, which quantifies how appropriately some micro-outcome o_k can be said to be supervened upon by macro-outcome B. In the previous example calculation, S^* takes only extreme values (0 or 1) for each micro-outcome, which would mean that every micro-outcome can be perfectly sorted under some unique macro-outcome. Alternatively, S^* may range from 0 to 1 (exclusive). In this case, the contribution of a micro-outcome to the probability of a macro-outcome would still be proportional to how appropriately the macro-outcome in question can be said to supervene upon the micro-outcome, but such supervenience relationships would come in degrees.[21] Another way to pose this solution would be to concede that o_k realizes some definite macrostate, and then interpret S^* as a similarity relation between that definite macrostate and the macrostate under consideration. It is not my present concern to weigh in any further on these various approaches. The point is just that we could adapt this formalism of macro-probabilities to accommodate such philosophical concerns, if needed.

6 Objective Modal Structure: Metaphysical Considerations

I have attempted to cut the CF-PIF free from several theoretical and ontological commitments that I see as unnecessary. The commitment to objective modal structures is not among those that can be cut free, due to the modal character of objective probability and causation. I now intend to hedge against metaphysically inflationary interpretations of this modal structure.

[21] An example of a view with which these intermediate values of S^* may be compatible can be found in the 'real patterns' view of Dennett (1991), which he describes as a kind of 'mild realism' with respect to macroscopic entities. This would also represent one way in which microstates could unequally share a macrostate's possibility space.

I have endorsed a view of probabilistic causation that invokes volumetric measures of modal structures (Lyon, 2011). At the same time, I have resisted a view that invokes dispositional powers as irreducible properties of entities responsible for generating modal structure (Triviño and Nuño de la Rosa, 2022). And, furthermore, I now seek to distance myself from other views that would likewise profess to know of the existence and nature of any such metaphysical entities like fundamental powers, which supposedly lie between what is empirically accountable or logically certain. These are not jointly inconsistent aims. Allow me to briefly explain why.

The most radically inflationary treatments of modal metaphysics are exemplified by those who would argue that there is a being who occupies all metaphysically possible worlds and thus could not have failed to exist (Rasmussen, 2016); or that all metaphysically possible worlds are concrete and actual worlds (Lewis, 1986). Inflationary approaches which hold that we should otherwise reify metaphysical modality as distinct from logical modality, untethered from empirical evidence, and knowable by the light of reason, have become the mainstream in analytic philosophy.

I follow in the footsteps of a number of philosophers of science who are dissatisfied with this mainstream inflationism and yet still recognize that we will need some sort of objective treatment of modality if we are to assert that any scientific theories correctly or incorrectly describe the ways the world would behave in counterfactual scenarios (see, inter alia, Ladyman and Ross 2007; Norton 2022; and Woodward 2023). Scientific theories have a modal character - they tell us what would and would not happen in counterfactual scenarios - and yet they are empirically accountable. I think we need no conception of non-logical modality beyond this empirically accountable sort; but even if one thinks we might need a more inflationary conception for some purposes, the purposes of the CF-PIF fit safely within the former's bounds. Of course, what this empirically accountable modality itself really is - e.g., the degrees of theoretical freedom allowed by a body of evidence (Norton, 2022), a primitive ontological commitment of theories which are themselves empirically accountable (Ladyman and Ross, 2007), or something else entirely - remains open for debate.

7 Conclusion

The conceptual foundation of the PIF holds that fitness is an objectively probabilistic causal disposition toward reproductive success. This interpretation affords the concept of fitness real explanatory purchase over the behavior of the biosphere. I have argued that the theoretical and ontological commitments of the PIF's conceptual foundation are less burdensome than has typically been recognized. I first detailed how objectively probabilistic causation constitutes the explanatory machinery of the PIF (see Section 2). I then argued that the PIF is (i) compatible with a principled pluralism regarding the mathematical relationship between measures of individual and trait reproductive success (see

Section 3); (ii) independent of the propensity interpretation of probability (see Section 4); and (iii) independent of microphysical indeterminism (see Section 5). It is ontologically committed to objective modal structures, but only in the minimal and empirically accountable sense required to say that any scientific theory correctly or incorrectly describes the ways the world would behave in counterfactual scenarios (see Section 6).

Many of the considerations discussed above, especially those in Sections 4, 5, and 6, are quite general and may have implications well beyond the domain of fitness explanations. A wide variety of explanatory frameworks could be built using the same basic explanatory machinery that figures in the PIF, namely, objectively probabilistic causation. For instance, Nuño de la Rosa and Villegas (2022) develop a view that interprets evolvability and variability as propensities. By clarifying the theoretical and ontological commitments of the PIF's conceptual foundation, my hope is that discussions of such interpretations of other scientific concepts may also benefit.

Acknowledgments. For countless hours of revision and discussion on earlier versions of this work, I owe a special thanks to Prasanta Bandyopadhyay. I am also grateful to many others for insightful comments, including Gordon Brittan Jr., Sahotra Sarkar, Caitlin Mace, John Norton, Alan Love, Joel Velasco, Scott Gilbert, Stuart Newman, Kate MacCord, Kristen Intemann, Bonnie Sheehey, Daniel Flory, and two anonymous referees. Their contributions have markedly improved this work, and any flaws that remain do so only in spite of their efforts.

Declarations

- **Funding** This work was presented as a poster at the Philosophy of Science Association (PSA) 2022 Meeting, the travel for which was funded by the PSA's associated NSF travel grant, as well as by the History and Philosophy Department of Montana State University.
- **Competing interests** The author has no competing interests to disclose.

References

Abrams, M. 2006. Infinite populations and counterfactual frequencies in evolutionary theory. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 37*(2): 256–268. https://doi.org/10.1016/j.shpsc.2006.03.004 .

Abrams, M. 2007. Fitness and propensity's annulment? *Biology and Philosophy 22*(1): 115–130. https://doi.org/10.1007/s10539-005-9010-x .

Abrams, M. 2009. The unity of fitness. *Philosophy of Science 76*(5): 750–761. https://doi.org/10.1086/605788 .

Abrams, M. 2012. Mechanistic probability. *Synthese 187*(2): 343–375. https://doi.org/10.1007/s11229-010-9830-3 .

Ariew, A. and Z. Ernst. 2009. What fitness can't be. *Erkenntnis 71*(3): 289–301. https://doi.org/10.1007/s10670-009-9183-9 .

Batterman, R.W. and C.C. Rice. 2014. Minimal model explanations. *Philosophy of Science 81*(3): 349–376. https://doi.org/10.1086/676677 .

Beatty, J. and S. Finsen. 1989. Rethinking the propensity interpretation: A peek inside pandora's box, In *What the Philosophy of Biology is*, ed. Ruse, M., 17–30. Springer Dordrecht.

Bigelow, J. 1976. Possible worlds foundations for probability. *Journal of Philosophical Logic 5*(3): 299–320. https://doi.org/10.2307/30226146 .

Bourrat, P. 2017. Explaining drift from a deterministic setting. *Biological Theory 12*(1): 27–38. https://doi.org/10.1007/s13752-016-0254-2 .

Boyd, R. 2017. How philosophers 'learn' from biology: Reductionist and antireductionist "lessons", In *How Biology Shapes Philosophy: New Foundations for Naturalism*, ed. Livingstone Smith, D., 276–301. Cambridge University Press.

Brandon, R.N. 1978. Adaptation and evolutionary theory. *Studies in History and Philosophy of Science Part A 9*(3): 181–206. https://doi.org/10.1016/0039-3681(78)90005-5 .

Brandon, R.N. 1990. *Adaptation and Environment.* Princeton University Press.

Clatterbuck, H., E. Sober, and R. Lewontin. 2013. Selection never dominates drift (nor vice versa). *Biology & Philosophy 28*(4): 577–592. https://doi.org/10.1007/s10539-013-9374-2 .

de Canson, C. 2022. Objectivity and the method of arbitrary functions. *The British Journal for Philosophy of Science 73*(3): 663–684. https://doi.org/10.1093/bjps/axaa001 .

Dennett, D.C. 1991. Real patterns. *The Journal of Philosophy 88*(1): 27–51. https://doi.org/10.2307/2027085 .

Drouet, I. and F. Merlin. 2015. The propensity interpretation of fitness and the propensity interpretation of probability. *Erkenntnis 80*(S3): 457–468. https://doi.org/10.1007/s10670-014-9681-2 .

Godfrey-Smith, P. 2007. Conditions for evolution by natural selection. *The Journal of Philosophy 104*(10): 489–516. https://doi.org/10.5840/jphil2007104103 .

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

Humphreys, P. 1985. Why propensities cannot be probabilities. *The Philosophical Review 94*(4): 557–570. https://doi.org/10.2307/2185246 .

Humphreys, P. 2004. Some considerations on conditional chances. *The British Journal for the Philosophy of Science* 55: 667–680. https://doi.org/10.1093/bjps/55.4.667 .

Ismael, J.T. 2009. Probability in deterministic physics. *The Journal of Philosophy 106*(2): 89–108. https://doi.org/10.5840/jphil2009106214 .

Ismael, J.T. 2011. A modest proposal about chance. *The Journal of Philosophy 108*(8): 416–442. https://doi.org/10.5840/jphil2011108822 .

Ladyman, J. and D. Ross. 2007. *Every Thing Must Go: Metaphysics Naturalized.* Oxford University Press.

Lewis, D. 1980. A subjectivist's guide to objective chance, In *IFS: Conditionals, Belief, Decision, Chance and Time*, eds. Harper, W.L., R. Stalnaker, and G. Pearce, 267–297. Springer Netherlands.

Lewis, D. 1986. *On the Plurality of Worlds.* Wiley-Blackwell.

List, C. and M. Pivato. 2015. Emergent chance. *The Philosophical Review 124*(1): 119–152. https://doi.org/10.1215/00318108-2812670 .

Lyon, A. 2011. Deterministic probability: neither chance nor credence. *Synthese 182*(3): 413–432. https://doi.org/10.1007/s11229-010-9750-2 .

Mills, S.K. and J.H. Beatty. 1979. The propensity interpretation of fitness. *Philosophy of Science 46*(2): 263–286. https://doi.org/https://doi.org/10.1086/288865 .

Millstein, R.L. 2002. Are random drift and natural selection conceptually distinct? *Biology and Philosophy 17*(1): 33–53. https://doi.org/10.1023/A:1012990800358 .

Millstein, R.L. 2003. Interpretations of probability in evolutionary theory. *Philosophy of Science 70*(5): 1317–1328. https://doi.org/10.1086/377410 .

Millstein, R.L., M. Dietrich, and R. Skipper. 2009. (Mis)interpreting mathematical models of drift: Drift as a physical process. *Philosophy and Theory in Biology* (1): 1–13. https://doi.org/10.3998/ptb.6959004.0001.002 .

Mumford, S. and R.L. Anjum. 2011. *Getting Causes from Powers.* Oxford University Press.

Niiniluoto, I. 1988. Probability, possibility, and plenitude, In *Probability and Causality*, ed. Fetzer, J., 91–108. D. Reidel Publishing Company.

Norton, J. 2003. Causation as folk science. *Philosopher's Imprint 3*(4): 1–22 .

Norton, J. 2021. *The Material Theory of Induction.* University of Calgary Press.

Norton, J. 2022. How to make possibility safe for empiricists, In *Rethinking the Concept of Law of Nature: Natural Order in the Light of Contemporary Science*, ed. Ben-Menahem, Y., 129–159. Springer Cham.

Nuño de la Rosa, L. and C. Villegas. 2022. Chances and propensities in evo-devo. *The British Journal for the Philosophy of Science 73*(2): 509–533. https://doi.org/10.1093/bjps/axz048 .

Otsuka, J. 2016. A critical review of the statisticalist debate. *Biology and Philosophy 31*(4): 459–482. https://doi.org/10.1007/s10539-016-9528-0 .

Pence, C. 2021. *The Causal Structure of Natural Selection.* Cambridge University Press.

Pence, C. and G. Ramsey. 2013. A new foundation for the propensity interpretation of fitness. *The British Journal for the Philosophy of Science* 63: 851–881. https://doi.org/10.1093/bjps/axx031 .

Pence, C. and G. Ramsey. 2015. Is organismic fitness at the basis of evolutionary theory? *Philosophy of Science 82*(5): 1081–1091 .

Poincaré, H. 1896. *Calcul des Probabilités.* Carré et Naud.

Popper, K.R. 1959. The propensity interpretation of probability. *The British Journal for the Philosophy of Science 10*(37): 25–42. https://doi.org/10.1093/bjps/x.37.25 .

Rasmussen, J. 2016. Could god fail to exist? *European Journal for Philosophy of Religion 8*(3): 159–177. https://doi.org/10.24204/ejpr.v8i3.1692 .

Richardson, R.C. 1996. Critical notice: Robert N. Brandon, *Adaptation and Environment. Philosophy of Science 63*(1): 122–136. https://doi.org/10.1086/289899 .

Richardson, R.C. and R.M. Burian. 1992. A defense of propensity interpretations of fitness. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (Volume One: Contributed Papers): 349–362. https://doi.org/10.1086/psaprocbienmeetp.1992.1.192767 .

Rosenberg, A. 1983. Fitness. *The Journal of Philosophy 80*(8): 457–473. https://doi.org/10.2307/2026163 .

Rosenthal, J. 2010. The natural-range conception of probability, In *Time, chance, and reduction: Philosophical aspects of statistical mechanics*, eds. Ernst, G. and A. Huttemann, 71–90. Cambridge University Press.

Salmon, W. 1979. Propensities: A discussion review. *Erkenntnis 14*(2): 183–216 .

Sarkar, S. 2011. Drift and the causes of evolution, In *Causality in the Sciences*, eds. Illari, P.M., F. Russo, and J. Williamson, 445. Oxford University Press.

Sober, E. 1984. *The Nature of Selection: Evolutionary Theory in Philosophical Focus.* University of Chicago Press.

Sober, E. 2001. The two faces of fitness, In *Thinking About Evolution: Historical, Philosophical, and Political Perspectives*, eds. Singh, R.S., C.B. Krimbas, D.B. Paul, and J. Beatty, Volume 2, 309–321. Cambridge University Press.

Sober, E. 2010. Evolutionary theory and the reality of macro-probabilities, In *The Place of Probability in Science: In Honor of Ellery Eells (1953-2006)*, eds. Eells, E. and J. Fetzer, Boston Studies in the Philosophy of Science, 133–161. Springer Netherlands.

Sober, E. 2013. Trait fitness is not a propensity, but fitness variation is. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences 44*(3): 336–341. https://doi.org/10.1016/j.shpsc.2013.03.002. .

Sober, E. 2020. Fitness and the twins. *Philosophy, Theory, and Practice in Biology 12*(1): 1–13. https://doi.org/10.3998/ptpbio.16039257.0012.001 .

Strevens, M. 2011. Probability out of determinism, In *Probabilities in Physics*, eds. Beisbart, C. and S. Hartmann, 339–364. Oxford University Press.

Suárez, M. 2013. Propensities and pragmatism. *The Journal of Philosophy 110*(2): 61–92. https://doi.org/10.5840/jphil2013110239 .

Suárez, M. 2022. The complex nexus of evolutionary fitness. *European Journal for Philosophy of Science 12*(1): 1–26. https://doi.org/10.1007/s13194-021-00434-w .

Triviño, V. and L. Nuño de la Rosa. 2022. A causal dispositional account of fitness. *History and Philosophy of the Life Sciences 38*(3): 1–18. https://doi.org/10.1007/s40656-016-0102-5 .

von Plato, J. 1983. The method of arbitrary functions. *The British Journal for the Philosophy of Science 34*(3): 37–47. https://doi.org/10.1093/bjps/34.1.37 .

Woodward, J. 2003. *Making Things Happen.* Oxford University Press.

Woodward, J. 2023. Sketch of some themes for a pragmatist philosophy of science, In *The Pragmatist Challenge: Pragmatist Metaphysics for Philosophy of Science*, eds. Andersen, H.K. and S.D. Mitchell, 15–66. Oxford University Press.