The Fine-Tuning Argument and the Requirement of Total Evidence

Peter Fisher Epstein*†

Abstract. According to the *Fine-Tuning Argument* (*FTA*), the existence of life in our universe confirms the *Multiverse Hypothesis* (H_M). A standard objection to FTA is that it violates the *Requirement of Total Evidence* (*RTE*). I argue that RTE should be rejected in favor of the *Predesignation Requirement*, according to which, in assessing the outcome of a probabilistic process, we should only use evidence characterizable in a manner available prior to observing the outcome. This produces the right verdicts in some simple cases in which RTE leads us astray; and, when applied to FTA, it shows that our evidence does confirm H_M.

*To contact the author, please write to: Peter Epstein, Department of Philosophy, 314 Moses Hall #2390, University of California, Berkeley, CA 94720-2390; e-mail: pepstein@berkeley.edu.

1. Introduction. According to the *Fine-Tuning Argument* (*FTA*), the existence of life in our universe confirms the hypothesis that there are many universes (the *Multiverse Hypothesis* (H_M)) over the hypothesis that our universe is the only one (the *Universe Hypothesis* (H_U)). Roger White (2000), following Ian Hacking (1987), has challenged FTA on the grounds that the reasoning behind it violates the *Requirement of Total Evidence (RTE)*, which dictates that we must always assess hypotheses in light of the *most specific* evidence at our disposal. Below, I show that, despite its near universal acceptance, RTE delivers the wrong verdict in cases like that involved in the debate about FTA. Such cases, I argue, ought to be assessed in light of a principle that is already embodied in sound scientific practice: the *Predesignation Requirement*, which demands that, in assessing the outcome of a probabilistic process, we characterize both the hypotheses we are evaluating and the outcome observed in a way that was salient prior to actually observing the outcome. Following the Predesignation Requirement means ignoring some of our evidence (namely, evidence characterized in a way *not* salient prior to observing the outcome); it thus means violating RTE. But, as I will show, following the Predesignation Requirement gives us the right verdicts in some simple cases in which RTE leads us astray; and, when applied to FTA, the Predesignation Requirement leads to the conclusion that our evidence does indeed confirm H_M.

2. White's Objection and RTE. The reasoning behind FTA seems, at first glance, quite straightforward: The conditions needed to support life are extremely unlikely to occur in any given universe, according to our best cosmological theories. So those conditions would be much more likely to occur if there were many universes. We know that those conditions have indeed occurred (just look around!), so we have evidence that H_M is true.

White's objection to the argument begins with the observation that we in fact have two pieces of evidence about the existence of a life-supporting universe: we know that *some* universe is life-supporting (call this evidence E_G); but we also know that *this specific* universe (*our* universe; call it α) is life-supporting (call this evidence E_S). E_G, on its own, does indeed support H_M over H_U: the chances of there being *some* life-supporting universe are higher the more universes there are. E_S, on the other hand, does not support H_M over H_U: α's chances of supporting life are insensitive to whether many (or any) other universes exist. Thus, in suggesting that our evidence makes H_M more likely than H_U, advocates of FTA must be relying on the less-specific E_G rather than the more-specific E_S; their reasoning should therefore be rejected, according to White, because it violates RTE.

White's analysis of FTA is based on a Bayesian framework for assessing competing hypotheses, according to which hypothesis H_1 is confirmed over hypothesis H_2 by evidence E just in case E is more likely to have occurred given H_1 than it is given H_2 (see Jeffrey et al. (1962), Hacking (1965), Sober (2009)). We can label this the *Likelihood Principle* (*LP*), and we can formalize it as follows:

1

LP. E is evidence for H_1 over H_2 iff $P(E|H_1) > P(E|H_2)$

So, in claiming that our evidence confirms H_M over H_U, advocates of FTA are claiming that we have some piece of evidence, E, such that $P(E|H_M) > P(E|H_U)$.[1] In order to evaluate this claim, we will need to determine what the relevant evidence (E) is. The two possibilities seem to be E_S and E_G.

As noted above, White contends that E_G confirms H_M over H_U, but E_S does not. To see why, think of the beginning of each universe as a causally-isolated chance event, say the spinning of a giant roulette wheel with a billion slots. The fine-tuning data suggest that a universe's being life-supporting is extremely unlikely; it is like hitting a single specific slot on the billion-slotted wheel.[2] So the chance of *some* universe's being life-supporting is greater if there are many universes, just as the chance of *some* roulette wheel's hitting that lucky number is greater if there are many wheels. This shows that E_G supports H_M over H_U.

But if we focus just on α, and *its* "roulette wheel," we can see that it is no more likely that α would be life-supporting if there are many other universes than if ours is the only one. According to H_M, each universe that exists on is causally isolated from the others; so whether universes other than α exist cannot affect the likelihood that α is life-supporting.

Summing this up in terms of LP, we have:

$$P(E_G|H_M) > P(E_G|H_U)$$
$$P(E_S|H_M) = P(E_S|H_U)$$

---

[1] Note that the claim at issue concerns whether the evidence about the existence of life offers *incremental confirmation* of H_M over H_U. In assessing this question, I will not be assessing whether it is *overall* rational to accept H_M. The latter question depends on many factors, including how well our total scientific evidence (including our knowledge of the specific values of constants in our universe) supports various specific cosmological theories. There are many different versions of H_M in the scientific literature, which are supported to varying degrees by contemporary physics. I won't be evaluating particular versions of H_M individually here.

[2] This analogy to randomizing mechanisms, like roulette wheels, in discussions of the origins of the universe is common, if somewhat obscure. The idea is that the values of the fundamental constants and the initial conditions have to be "set" within an incredibly narrow range to even allow for matter to exist, so a given big bang's resulting in life-supporting conditions is incredibly improbable. But the conditions are of course not "set" by any *spatiotemporal process* that intelligibly mirrors the spinning of a giant roulette wheel; so it is unclear how we ought to think of assigning probabilities in this context. I'll follow White in ignoring this complication here.

The crucial question now becomes which piece of evidence—E_G or E_S—is the one we should use in assessing the likelihood of H_M and H_U.

White (2000, 264) claims that E_S is the version of the evidence we must use in assessing H_M and H_U. He defends this claim by pointing to RTE, which he states as a general principle of reasoning: "in the confirming of hypotheses, we cannot, as a general rule, set aside a specific piece of evidence in favor of a weaker piece." E_S is a more specific version of E_G. RTE thus requires that we use E_S rather than E_G in evaluating hypotheses; thus, we do not get any confirmation of H_M over H_U.

RTE plays a crucial role in White's argument against FTA, and, intuitively, the principle has some appeal: it seems reasonable to think that we should always use *all* of the information at our disposal in assessing our evidence. But, as I will show below, there are cases where RTE should *not* be applied: in some cases, proper reasoning dictates that we use a more-general version of the evidence at hand, rather than a more-specific version.

3. Trouble for RTE. Consider the following case, introduced by White (2000, 267):

> **Case B.** Jane knows that an unspecified number of players will simultaneously roll a pair of dice just once, and that she will be woken if, and only if, a double-six is rolled. Upon being woken she infers that there were several players rolling dice.

White's (plausible) assessment of this case is that Jane's reasoning is cogent. The evidence she obtains upon being woken is:

> **E_G.** Some player rolled a double-six.

And the hypothesis that there were multiple players (H_{MP}) is confirmed over the hypothesis that there was just one player (H_{OP}) by E_G, since $P(E_G|H_{MP}) > P(E_G|H_{OP})$. That is, it is more likely that some player rolled a double-six if there were many players rather than just one.

Here, White condones Jane's use of a *general* statement of her evidence ("Some player rolled a double-six") in assessing the likelihood of the hypotheses, so we might wonder whether White's own case is a counter-example to RTE. According to White, though, Jane does not violate RTE because she *is* using the most-specific version of her evidence; White claims that Jane simply *lacks* any more-specific evidence: "Jane's evidence is not about the outcome of a particular roll, but simply the fact that she has been woken" (White 2000, 268).

This claim is, however, simply false. Jane *can* state her evidence in the general form "Some player rolled a double-six," but she could also state it in a more specific form: "*The player whose roll actually led to my being awoken* rolled a double-six." Jane can even use this description to *rigidly designate* that player, perhaps inventing a name for

3

that player (call her "Asha"), *a la* Kripke's story about Jack the Ripper.[3] So Jane also has the following piece of evidence:

E_S. Asha rolled a double-six.

Does E_S confirm H_{MP}? That is, was it more likely that *Asha's* roll would come up double-six if there were many players? Certainly not: no matter how many other players there were (zero or a million), Asha's roll had a 1/36 chance of being double-six. So $P(E_S|H_{MP})$ = $P(E_G|H_{OP})$.

We now have a case with the structure described in §1: a more-specific characterization of the evidence, E_S, does not confirm H_{MP} over H_{OP}, while a less-specific characterization, E_G, does. And, if we follow RTE, we will conclude that Jane, in spite of initial appearances, has no reason to favor H_{MP} over H_{OP}. Just by going through a bit of clever rigid-designation, Jane has vitiated the force of her evidence!

This is surely the wrong verdict; and so, we seem to have found a case in which proper reasoning dictates that we violate RTE, and stick to a less-specific formulation of our evidence in applying LP. Still, we might think this is a rather limited result: the clever rigid-designation involved in the case might seem fishy, and the dice-rolling scenario itself is rather artificial. But a further case reveals just how far-reaching a threat RTE, taken as a fully general principle, would pose to our ability to confirm hypotheses in the light of our evidence. Consider the following example:[4]

> **Fish.** You know there are some large fish and some small fish in a lake, but you don't know the proportions: either there are mostly large fish (H_L) or mostly small fish (H_S). You know that you will catch one fish each time you stick in your net. You stick in your net and catch a large fish.

Let's describe your evidence in the following way:

E_G. You caught some large fish.

E_G would seem to give you evidence in favor of H_L over H_S, since you are more likely to catch a large fish if the majority of the fish are large than if the majority of the fish are

---

[3] The label "Jack the Ripper" was coined to denote the person killing people in London in 1888; the reference of the term was initially fixed through that description. But the term, having had its reference fixed in this way, could then function as a rigid designator for a particular person: we could say things like, "Jack the Ripper might not have gone around killing those poor women if he'd only had a happier childhood" (see Kripke 1980, 79).

[4] This example, originally due to Eddington (1939), is adapted from Kotzen (2012).

small: $P(E_G|H_L) > P(E_G|H_S)$. Reasoning in this way seems like a paradigm example of empirically confirming a hypothesis, and, as such, it seems beyond reproach.

But wait! The fish you caught, in addition to being large, is a *particular* fish (as fish tend to be). You name her Asha (ostensively, without need of any clever rigid-designation). So you also have the evidence:

E$_S$. You caught Asha.

We can now ask whether E$_S$ is more likely given H$_L$ than given H$_S$: Were you more likely to catch *Asha* if there were mostly large fish, rather than mostly small fish?

It seems clear that you were equally likely to catch Asha on either hypothesis: whatever the size of the *other* fish, the likelihood that your net would happen upon the specific large fish Asha was exactly the same. Asha was swimming along, and happened to be in the wrong place at the wrong time. The probability of that event is presumably sensitive to such factors as how frequently Asha swims in that area of the lake. But one thing that seems completely irrelevant is the size of the other fish around her. So, once again, the specific evidence fails to confirm the hypothesis: $P(E_S|H_L) = P(E_S|H_S)$. Thus, if we follow RTE, we will be led to the false conclusion that we get no confirmation of H$_L$ over H$_S$.

The above cases show that we should not always follow RTE in evaluating the likelihood of competing hypotheses. There are cases in which proper reasoning requires ignoring a more-specific statement of our evidence in favor of a less-specific version. Thus, White cannot simply appeal to RTE to justify his claim that we must reason on the basis of E$_S$, rather than E$_G$, in assessing FTA.

We might wonder, though, whether assessing the probability of H$_M$ in the context of FTA is a case like the ones I've been describing, in which we should use our less-specific evidence. For there certainly are other cases in which we should use our more-specific evidence. Consider another case of White's:

Case B*. Jane knows that she is one of an unspecified number of sleepers each of whom has a unique partner who will roll a pair of dice. Each sleeper will be woken if and only if her partner rolls a double-six. Upon being woken, Jane infers that there are several sleepers and dice-rollers. (White 2000, 268)

Jane knows a more-specific and a less-specific fact:

E$_G$. Some player rolled double-six.
E$_S$. *Jane's* partner (call her "Asha") rolled double-six.

And we again have the same evidential structure: The more-specific evidence does not confirm the hypothesis, since the probability that *Asha* would roll double-six is insensi-

tive to the existence of other sleepers and other rollers ($P(E|H_{MP}) = P(E|H_{OP})$). But the less-specific evidence *does* confirm the hypothesis: it is more likely that *some* player will roll double-six if there are many players ($P(E|H_{MP}) > P(E|H_{OP})$). And in *this* case, it seems Jane would do well to focus on her more-specific evidence, E_S: Jane doesn't have any reason to think there were other sleepers when she wakes up, because she knew that either way (whether she was the only sleeper or one of many), *her* chances of being awoken were simply 1/36. Imagine Jane the night before she goes to sleep, wondering whether there are other sleepers, and trying to comfort herself by thinking there are. This is a badly misguided attempt at self-help: the presence of compatriots—though it makes it more likely that *some* sleeper will be awoken—fails to improve Jane's own chances of waking up, which depend solely on Asha's roll.

So what marks off the cases in which we should use our more specific evidence (Case B*) from those in which we should not (Case B; Fish), and on which side of the line does the FTA evidence fall? The situation here is quite puzzling: cases that seem structurally identical require different applications of LP; and, in some of those cases, we seem to be required to ignore parts of what we observe in order to correctly assess the evidential import of our observations. In what follows, I will argue that, in order to sort out these cases, we must address two distinct instances of the *reference class problem* that arise in connection with FTA: in applying LP to the evidence at hand, we need to attend both to (1) how we classify the contents of our observations; and (2) how we think of our own role as observers.

4. The Reference Class Problem. Assessing the evidential import of an observation using LP requires assigning that observation a probability, conditional on two different hypotheses. As a general rule, in order to assign probabilities to observations, those observations must be classified as members of a class of *possible outcomes*. Suppose you roll a die and get a six. Someone asks, "How likely was that outcome?" Before answering, you will need to *classify* the outcome in some way. If you classify it as a member of the class of *sixes*, the answer is 1/6, since one-sixth of the possible outcomes of a die-roll fall into that class. If you instead classify the very same outcome as a member of the class of *evens*, the answer is ½. The different answers here turn on which of the possible outcomes that do *not* occur—a four, for example—should be counted as an instance of the same *type* as the actual outcome. And that depends on which type, out of the many under which it falls, you use in classifying the observed outcome (the six you actually rolled).

This question—Which type is relevant in classifying the outcome?—is what is at issue in the cases discussed above. In applying LP to determine whether we get confirmation from our evidence, we need to decide between two different descriptions of an observed outcome, which designate that outcome in terms of its membership in two different classes of possible outcomes. E_G describes the outcome as falling under a broad class of possible outcomes (in Case B*, for instance, E_G classifies the outcome as a member of

the class of *double-six rolls*), while E_S describes it as falling under a narrower sub-class (E_S classifies that same outcome as a member of the class of *double-six rolls by Asha*). What we are interested in, in applying LP, is not the unconditional probability of the observed outcome (P(E)), but rather the conditional probabilities $P(E|H_1)$ and $P(E|H_2)$. And, as we've seen, those conditional probabilities (which determine whether the observation confirms H_1 over H_2) change, depending on which reference class we use in stating the evidence E. Thus, what we have is a classic example of the reference class problem, as described by Alan Hájek: "The reference class problem arises when we want to assign a probability to a single proposition [or event], *X*, which may be classified in various ways, yet its probability can change depending on how it is classified.… *X* may be classified as belonging to set S_1, or to set S_2…. *Qua* member of S_1, its probability is p_1; *qua* member of S_2, its probability is p_2, where $p_1 \neq p_2$" (Hájek 2007, 565).

So we could decide which version of the evidence to use in each case if we had a solution to the reference class problem – if we had some rule for deciding which reference class was relevant in assessing the conditional probabilities associated with a given observed outcome. One way of understanding RTE is as proposing just such a rule: according to RTE, we should always use the *narrowest* reference class under which we can classify our observation in assessing that observation's evidential import. But, as we saw above, RTE sometimes fails to deliver the correct verdict, and so it must be rejected.

I won't be offering an alternative general solution to the reference class problem. Instead, I want to draw out some lessons about how to deal with the problem in the particular kinds of cases discussed above. What those cases reveal, I claim, is that, in applying LP, we should describe the observed outcome in terms of a reference class of possible outcomes that was salient *antecedent to* the observation itself. That is, in assessing the evidential import of the outcome of some probabilistic process, we must not characterize that outcome in terms of a reference class that became salient only because of the specific result of the probabilistic process. Such *post hoc* classifications of the possible outcomes of the process can mask the genuine confirmation an observation provides.

5. The Predesignation Requirement. Consider the case of Fish, in which the outcome event is the catching of a specific large fish, Asha. There are two relevant reference classes under which we could classify this outcome: *large-fish-catchings* and *Asha-catchings*. These different choices of reference class give us two different partitions of the space of possible outcomes: we can divide the possible outcomes into Asha-catchings and Non-Asha-Catchings, or into Large-Fish-Catchings and Small-Fish-Catchings (see figure 1).

As argued above, if we employ LP by using the more specific reference class, we get the wrong result: since the observation, *characterized as a member of the class of Asha-catchings*, is no more likely on H_L than on H_S (as shown in the right-hand side of figure 1), employing LP in this way would lead us to conclude (falsely) that we have no confirmation of H_L over H_S.
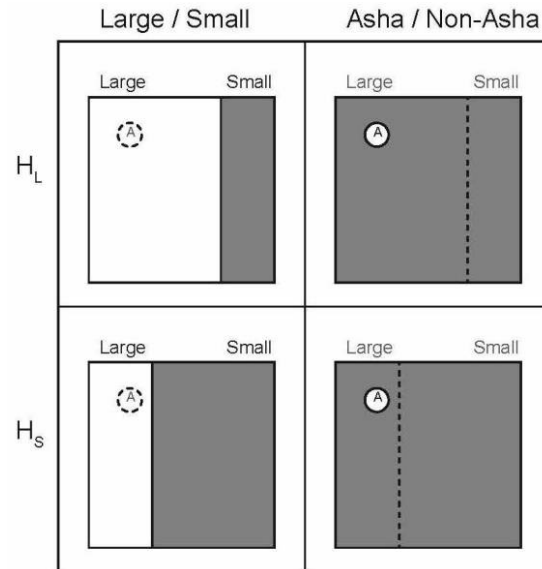
Figure 1. The space of possible outcomes is partitioned in two different ways (left-hand side vs. right-hand side), as determined by the two relevant reference classes into which the actual outcome falls. Under each partition, the outcomes that are in the same reference class as the actual outcome form the unshaded region; those that are not form the shaded region. (Equivalently: the size of the unshaded regions in the left-hand side of the diagram represent the conditional probabilities $E_G|H_L$ and $E_G|H_S$, while those in the right-hand side represent the conditional probabilities $E_S|H_L$ and $E_S|H_S$.) Comparing the relative sizes of the unshaded regions in the upper and lower rows under each of the partitions, we see that whether the likelihood of the observed outcome is greater given H_L than it is given H_S (whether the unshaded region is larger in the top row than in the bottom row) depends on whether we classify that outcome as a *large-fish-catching* (left-hand side) or an *Asha-catching* (right-hand side).

Intuitively, the reason we get the wrong result here is that the more specific reference class does not accurately capture the nature of the experiment performed. In sticking our net into the lake, we didn't antecedently partition the possible outcomes into Asha-catchings and non-Asha-catchings, and hope to obtain evidence about the fish population in virtue of our observation falling into one of those two classes. We didn't have a view about how each of the possible "Asha outcomes" would affect our credence in the competing hypotheses, because we simply didn't have any way of thinking of Asha at all. Instead, we came to be focused on the Asha-catchings reference class only as a *result* of the actual outcome of the experiment. So to treat the result of the experiment as providing evidence in virtue of falling into the class of Asha-catchings is to misrepresent the evidential import of the probabilistic process that we observed.

The problem here stems from a failure to respect what is sometimes called the "Pre-designation Requirement," a widely-accepted rule of sound scientific practice (see Staley

8

2002). It is a familiar point that experimental design should ideally have the following structure. First, a hypothesis is generated. Then, an experimental procedure is designed to test the hypothesis against an alternate hypothesis (often the null hypothesis). The possible outcomes of the procedure are identified, and the evidential import of each possible outcome is assessed: the experimenters determine how their credences in the competing hypotheses ought to be updated, if a given possible outcome is observed. Then the experiment is run, the observation is made, and credences are updated accordingly.

A standard worry in experimental science arises when, instead of following this procedure, experimenters first observe the outcome of the experiment, and then use that outcome to generate hypotheses *post hoc*. Notoriously, such violations of the Predesignation Requirement—sometimes referred to as "data-mining" or "over-fitting"—can lead to the appearance of statistically significant experimental outcomes when no real confirmation has occurred.

When we use the narrower reference class in describing our evidence in the case of Fish we are guilty of a similar, but distinct, violation of the Predesignation Requirement. In this case, the *hypotheses* being tested—H_L and H_S—are not generated on the basis of the experimental outcomes. Instead, what is determined *post hoc*, on the basis of the observed outcome, is the partition of the possible outcome space: the Asha-catchings reference class is used to describe the observation, even though that reference class was entirely unknown prior to the observation of the outcome.

LP tells us that, in order to evaluate the evidential import of an observation E for hypothesis H, we should assess the conditional probability P(E|H). In familiar cases of Predesignation violation, the choice of H is made *post hoc*. This biases the results towards finding confirmation when there is none. In the cases I have been considering where we ought not use the more specific statement of our evidence, like Fish, it is instead the possible values of E that are identified *post hoc*. Here, we do not get an appearance of confirmation when there is none. Instead, by "over-fitting" the choice of *reference class* used in describing the possible outcomes to match the observed result, we mask the confirmation that the observation *does* provide.

One way to understand what goes wrong here is as follows. We caught a large fish, and, noting a particular feature of that fish about which we antecedently had no probabilistic beliefs—that the large fish was *Asha*—we used that particular feature to define a space of possible outcomes of the experiment. This was a mistake, since, so far as the experiment was concerned, Asha was not special: no matter what the outcome of the probabilistic process, we were going to catch *some* fish, and that fish was going to have some particular features that could be used to partition the space of possible outcomes. Suppose we had caught a *different* large fish, Byram. We would then have generated a *different* partition of possible outcomes: we would have assessed the outcome as a member of the class of *Byram-catchings*. If we partition the possible outcomes in such a way that the partition would be different for each different outcome, we mask the confirmation we can get from the observation by fitting it into a partition that we are prepared

9

to apply *uniformly* across the various possible outcomes: namely, the partition generated by classifying the outcome as a large-fish-catching, a partition about which we *did* have antecedent conditional credences. Using the narrower reference class in effect means that each possible outcome of the experiment will be utilized in a different way, thereby vitiating the force of the evidence a probabilistic experiment generates by fitting an observed outcome into a larger space of possible outcomes. Respecting the Predesignation Requirement ensures that we bring out an observation's evidential import, as the actual outcome of a probabilistic process that falls into an antecedently well-defined space of possible outcomes.

6. The FTA: A Second Reference Class Problem. I now turn to the case of FTA. In assessing whether our evidence confirms H_M over H_U, we are assessing the evidential import of the fact that the actual outcome of the "world-creation process" was a state in which α is life-sustaining. This outcome can be classified as a member of the class of possible states in which *some universe is life-sustaining* or as a member of the narrower class in which α *is life-sustaining*. The question to ask, given the above analysis, is whether the narrower reference class was salient prior to the actual outcome of the process.

It seems clear that it was not: we come to be acquainted with α, and to be in a position to know that it, in particular, is life-sustaining, only *because* α's big bang, in particular, resulted in a life-sustaining universe. So classifying our observation using a partition of the possible outcomes that makes reference to α fails to respect the Predesignation Requirement. It amounts to fitting our choice of reference class to the actual outcome of the probabilistic process, which, as we saw in the Fish case, leads to the false conclusion that we have no confirmation when in fact we do.

The story does not end here, however. For White, in a postscript to his paper, claims that we get no confirmation of H_M over H_U, even if we use the more general statement of our evidence, E_G. He suggests that the following principle dictates how we are to evaluate hypotheses in the light of observation:

> **Observation Principle.** An observation I make gives me evidence for hypothesis H only if it is more likely given H that I would make that observation. (White 2003, 244)

What I have been emphasizing is that, in applying such a principle, we need to be careful about how we characterize the observation made – how we classify the *content* of the observation. But White's point in framing the Observation Principle as he does is that an observation involves not only *what* is observed, but also the *subject* doing the observing. And, White claims, in the FTA case, even if I characterize the content of my observation in terms of the broader reference class *some life-sustaining universe exists*, I get no confirmation of H_M over H_U from my evidence, if I attend to the fact that *I* am the subject

10

who makes the observation. The reason is that, even though it is more likely that there is some life-sustaining universe given H_M, it is no more likely that *I would observe* such a universe. Since I couldn't exist in some other universe, the existence of a non-*α* life-sustaining universe is not something I could observe. So positing non-*α* universes, some of which are likely to be life-sustaining, doesn't raise the probability of *my observing* such a universe. Thus, the observation principle tells us that we get no confirmation of H_M over H_U from *either* E_S or E_G.

But just as the reference class problem arises when we consider how to characterize the content of the observation, so too does it arise when we consider how to characterize the observer. The outcome we know to obtain is that *I observe that α is life-sustaining*. As noted above, the content of this observation can be characterized by way of a broader reference class: *I observe that* some *universe is life-sustaining*. But we can also characterize the *observer* by way of a broader reference class: my observation is a member of the reference class *someone observes that some universe is life-sustaining*. Both aspects of the actual observation—its content and the observer—need to be classified in some way in order to assess the observation's evidential import.

This second reference class problem is crucial in assessing FTA. Consider figure 2. The actual outcome of the world-creation process—*I observe that α is life-sustaining*—falls into the circular region labeled "I-α" in each square. In the upper row of the diagram, the dotted circular region labeled "x-β" represents a possible outcome of the world-creation process on which *α* is not life-sustaining, and so *I* never exist (and, a fortiori, I never observe a life-sustaining universe); but another universe, *β*, is life-sustaining, and is observed to be so by some other observer. Such an outcome is only possible on H_M. Now, if we partition the space of possible outcomes based on whether *someone* observes a life-sustaining universe (left-hand side of the diagram), outcomes like x-β (those that fall outside of the I-α circle, but to the left of the vertical line, in the upper-left square) count as outcomes of the same type as the actual outcome. If we instead partition the possible outcomes based on whether *I* observe a life-sustaining universe (right-hand side of the diagram), outcomes like x-β do not count as outcomes of the same type as the actual observed outcome, and so the possibility of such outcomes does not make it more likely on H_M (as compared to H_U) that an outcome of the same type as the actual outcome would occur (the unshaded regions of the two squares on the right-hand side of the diagram are of equal area – they are simply the outcomes on which *I* observe a life-sustaining universe, namely *α*).

Thus, whether my observation confirms H_M over H_U turns on whether I should classify my evidence as falling into the reference class *observation of life-sustaining universe by me* or rather into the broader class *observation of life-sustaining universe by someone*. We need to address the reference class problem for the observer (for an extended discussion of this issue, see Bostrom 2002).
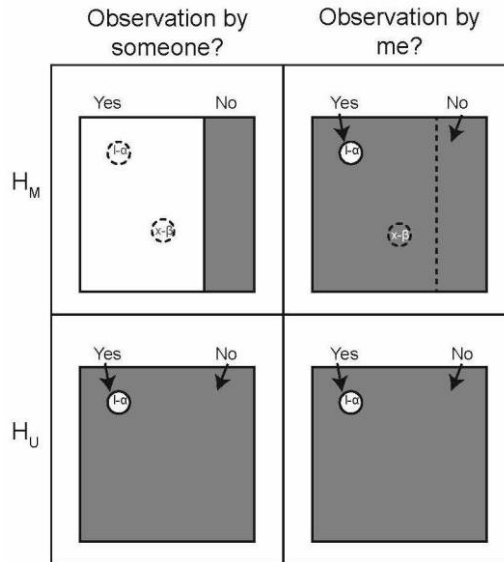
Figure 2. The space of possible outcomes of the "world-creation process" is partitioned in two different ways: on the left-hand side, outcomes are partitioned according to whether an observation of a life-sustaining universe is made by someone; on the right-hand side, outcomes are partitioned according to whether an observation of a life-sustaining universe is made by me. The actual outcome (where I observe that α is life-sustaining) is represented by the circle labeled "I-α". In the upper portion of the diagram, the circle labeled x-β represents another possible outcome, in which α is not life-sustaining, but a different possible universe, β, is observed to be life-sustaining by someone else, x. Whether the likelihood of the observed outcome is greater given H_M than it is given H_U (whether the unshaded region is larger in the top row than in the bottom row) depends on which partition (left side vs. right side) we use in classifying the observation.

White suggests that we answer the question of how to characterize the observer in the same way RTE advises us to answer the parallel question about how to characterize the content of the observation: according to White's Observation Principle, an observer ought to use the narrowest reference class into which she falls—she ought to classify her observation as an observation made specifically by *her*, as a unique individual—in assessing that observation's evidential import. But given the problems we saw above for RTE, we have reason to be suspicious of the parallel solution to the reference class problem when it comes to the observer. Instead of White's Observation Principle, I suggest that we approach the question of how to characterize the observer in the same way we approached the parallel question of how to characterize the content of the observation: we should ask whether using the narrower reference class violates the Predesignation Requirement.

12

The question thus becomes: In the case of our observation that *α* is life-sustaining, was the narrower reference class of *observers* a salient way to partition the space of possible outcomes antecedent to the actual observation of the outcome? Again, it seems clear that the answer is no. The partition of the space of possible outcomes of the world-creation process determined by my status as the observer of the outcome was no more salient prior to the observation of the actual outcome of that process than was the partition determined by focusing specifically on *α* as part of the content. I can classify the possible outcomes of the world-creation event as those in which *I* observe a life-sustaining universe and those in which I don't only *post hoc* – only because the specific outcome of the process (*α*'s turning out to be life-sustaining, which allows for *me* to exist) led me to focus on myself as observer. So a reference class determined by my own status as observer is an inappropriate one to use in assessing the evidential import of my observation. As seen in the case of Fish, we shouldn't characterize the outcome of a probabilistic process using a reference class that is salient only because of the specific outcome of the process. Doing so mischaracterizes the nature of the probabilistic process and its evidential connection to the hypotheses under consideration. By over-fitting our characterization of the outcome to the observation made, we mask the confirmation the outcome genuinely confers.

We might wonder, though, whether it makes sense to extend the Predesignation Requirement to the FTA case in this way. Recall that, in spelling out the rationale behind the Predesignation Requirement, I emphasized that violating the Requirement—by using a reference class that became salient only as a result of the actual outcome—would lead us to use *different* reference classes given different possible outcomes. In the case of Fish, for instance, we are poised to record whichever outcome occurs – whether Asha, or Byram, or some other fish is caught. As a consequence, using the identity of the particular fish we catch to generate a partition means we'll partition differently in different scenarios.

But in the FTA case, we would not partition the possible outcomes differently if the outcome were different—if, for example, *β* had been life-sustaining, while *α* was not— because we would not even *exist* in such a scenario. As White (2000, 268) notes, "it is not as though we were disembodied spirits, waiting for some big bang to produce a universe that could accommodate us." In the FTA case, because our identities as observers are tied to the particular universe *α*, we can only ever observe one kind of outcome—an outcome where *α* is life-sustaining—so there is no threat of our generating problematically gerrymandered partitions in each possible outcome scenario. And so, a defender of the Observation Principle might think, we should use the narrower reference class for the observer in the case of FTA, since doing so will not lead to different partitions in different outcome scenarios.

The problem with this suggestion is that it in effect treats our observation as an outcome of a process we were observing, where we were focused, in particular, on whether we would come to exist in *α*. It suggests that the relevant partition of possible outcomes divides those outcomes according to whether or not, as observers of the process in

question, we observe α—our predesignated potential "home universe"—to be life-sustaining. But such a picture is no more plausible than the one White mocks: just as "it is not as though we were disembodied spirits, waiting for some big bang to produce a universe that could accommodate us," it also is not as though we were disembodied spirits, keenly observing α—our designated potential home—and hoping that it, in particular, would be able to accommodate us. When considering the probabilistic process that led to our existence as observers, we cannot accurately capture the nature of our evidence by modeling that process as one that we—as the particular beings we are, tied specifically to α—observed. We don't have the proper perspective on such a process to assess its outcomes in terms of the observations we, specifically, could make.

The nature of the world-creation "process" itself is somewhat mysterious. As noted above (fn. 2), it is common in discussions of H_M to think of each big bang in the multiverse as an isolated chance event, like the spinning of a roulette wheel. This is what allows us to apply LP—our principle for assessing the evidential import of our observations—to hypotheses about the world-creation process. In applying LP, then, it seems we would be well-advised to map the unfamiliar scenario—one where we don't have a firm grip on the nature of the probabilistic "process," and one where our own identities as observers are not set in advance—onto a more familiar experimental paradigm. So we should not focus on the fact that *we* couldn't make observations of different outcomes, had they occurred; we should consider how an observer who *could* make such observations—an observer who could be seen as the "experimenter" trying to get evidence from the results of the probabilistic world-creation process—should assess the outcome observed. What would the outcome we know to have occurred—our observing α to be life-sustaining—tell an observer positioned to use such an observation, in virtue of standing to the world-creation process as we normally stand to our experiments?

Looked at in this way, the fact that the actual outcome is one in which *we* observe a life-sustaining universe seems irrelevant to the evidential import of our observation. From a God's-eye perspective, we can see that *whoever* came to exist—in whichever universe happened to be life-sustaining—could use her own specific identity to partition the space of possible outcomes. But doing so would lead those observers, taken as a whole, to carve up the possibility space in different ways, depending on the actual outcome of the process. And so, using a uniform partition of the possibility space that was salient *before* the outcome was observed—a partition according to whether *someone* observes a life-sustaining universe—more accurately captures the relevant information contained in the observation.

It is helpful to compare this analysis of FTA to Case B*. There, we saw that Jane ought not think she has confirmation of H_{MP} over H_{OP} based on her observation. But we might wonder why not. After all, the outcome in Jane's case could be characterized as *someone observes that someone rolled double-six*, rather than *Jane observes that someone rolled double-six*. And the likelihood of the former type of outcome is surely much higher on H_{MP} than on H_{OP}. For if H_{MP} is true, there will be worlds in which Jane does

not observe any outcome (because she never wakes up), but someone else observes a double-six. In Case B*, then, we will get the wrong result—we will falsely conclude that Jane gets confirmation of H_{MP} over H_{OP}—if we use the more general reference class in characterizing the identity of the observer.

The reason we get the wrong result if we fail to use the more specific characterization of the observer in Case B* is that, unlike in the FTA case, the partition of possible outcomes determined by Jane's specific identity as the observer was *antecedently* salient to Jane. She could antecedently consider the possible outcomes and partition them according to what she, in particular, would observe in each case. Such a partition of the space of possible outcomes does not rely on Jane's knowing the actual outcome itself, since Jane is antecedently acquainted with herself as a unique observer, and, knowing the set-up of the case, she knows that the observations she makes are tied to her specific partner's roll. So, unlike in the FTA case, partitioning the possible outcomes based on the specific identity of Jane *qua* observer does not involve any violation of the Predesignation Requirement; and it does not result in different partitions given different outcomes. Thus Jane, who can legitimately take herself to be the observer of the relevant probabilistic process (rather than just the particular *outcome* of it), ought to make use of the additional information that it was she, specifically, who made the observation.

The evidence I have in the FTA case, by contrast—my observation that α is life-sustaining—ought to be characterized as a member of a reference class that is broad along two dimensions. First, the *content* of the observation ought to be characterized as an outcome on which *some* universe, rather than specifically α, is life-sustaining, since partitions of the possible outcomes that make reference to α become salient only *post hoc*, as a result of the actual outcome of the world-creation process. Second, the observation ought to be characterized merely as an observation made by *some observer*, rather than specifically by me, since partitions that make reference to *me* become salient only as a result of the outcome. Putting these two claims together, the statement of the evidence that we should use in assessing the likelihood of H_M and H_U is: someone observes that some universe is life-sustaining. Since *that* outcome was more likely to occur given H_M than given H_U, we have confirmation of H_M over H_U.

7. Conclusion. My analysis has revealed that the issues surrounding FTA center on two distinct reference class problems that arise in evaluating H_M. Importantly, one of these reference class problems—the question of how to classify the *content* of an observation—arises independent of any of the issues about self-locating belief and selection bias that have been central to much of the debate about FTA (see, e.g., Bostrom 2002; Bradley 2009, 2012; Leslie 1988). For consider again the case of Fish. There, our evidence is not self-locating: it simply concerns the catching of a fish. And the selection procedure is unbiased: the net used is equally capable of catching any fish in the lake. But even in that case, we saw that we get the wrong result if we classify the outcome as specifically an *Asha*-catching, rather than more generally as a large-fish-catching. So neither information

15

about the selection procedure, nor the issue of self-locating belief, is central to the problem. What we need is a way to rule out the use of overly-specific evidence, even in mundane cases like Fish.[5] And the way to do that is to take note of how one comes to be focused on the more-specific reference class in question. We need to respect the Predesignation Requirement when deciding which partition of the possible outcomes we use in assessing the evidence: if a more specific reference class becomes salient only as a result of the actual outcome of a probabilistic process, that reference class is not an appropriate one to use in characterizing the evidence.

Draper et al. (2007) and Juhl (2005) offer a different kind of analysis of cases like Fish, one that does not involve the rejection of RTE. They both claim that the *specific* evidence, E_S (i.e., that we caught Asha), *does* confirm H_L over H_S.[6] In defending this claim, they argue that, while H_L makes it no more likely that we would catch Asha *given that she's in the lake*, it *does* make it more likely that Asha would *be in the lake* to be caught in the first place. And, if that were true, evaluating our evidence under the Asha-catchings partition would give us confirmation of H_L over H_S, since the increased likelihood of Asha's presence in the lake on H_L means that we are more likely to catch Asha on H_L. So, the claim goes, we are free to reason on the basis of the more specific evidence E_S, rather than E_G. We'll get the (correct) verdict that our evidence confirms H_L over H_S either way. Thus, the suggestion is that we need not abandon RTE in order to accommodate the Fish case.

The above line of thought relies on the claim that there is a probabilistic link between H_L and Asha's being in the lake: according to the argument just rehearsed, Asha is more likely to be in the lake if the lake contains mostly large fish. But there is simply no reason to believe in such a link. Since Asha's essence is, presumably, tied to her specific genetic origins, it's not as if *she* could have been born to different large-fish parents. So the presence of any large fish not in her direct lineage does not raise the probability of *her* being in the lake.

---

[5] This suggests that the analysis offered by Bradley (2009, 2012), which focuses on the role of bias in the selection procedure in FTA, will not on its own solve the problems generated by violations of the Predesignation Requirement. Bradley's discussion is directed at cases with a different structure from that of Fish; it is possible that his analysis could be extended in some way to cover such cases. But it is unclear how attending to the *un*biased selection procedure in such cases will solve the problem. Bradley (2009) does suggest one way we might still get confirmation in the FTA case, even if we assume there is no selection bias; his suggestion mirrors that made by Draper et al. and Juhl (see below).

[6] The cases discussed by Draper, et al. and Juhl do not involve populations of fish in a lake; but, since the cases are structurally analogous, I have framed these authors' claims in terms of the Fish example, for ease of exposition. I will note how the points about the simpler Fish case translate to the FTA case along the way.

Here is one way we might try to justify the claim that there is a probabilistic link between H_L and Asha's presence in the lake. Suppose the lake was populated as follows. First, God decided on the "demographics" He wanted for the lake: either it would be a lake of mostly large fish, or a lake of mostly small fish. Then, God selected (through a random process) the appropriate number of each kind of fish to make actual, from His stock of *possible large fish* and *possible small fish*. Since Asha was a possible large fish, she had a greater chance of being selected for "actualization" if there were more large fish "slots" available. Thus, Asha was more likely to be in the lake given H_L than given H_S.

But we of course have no reason at all to believe in such a story. Its metaphysical presuppositions are highly dubious, and our credence in its truth is essentially zero. So the claim that there is a probabilistic link between H_L and Asha's being in the lake, *given the way we actually think organisms come into being* (i.e., given that a fish's presence in a lake is sensitive only to her direct lineage's presence in that lake), should be rejected, as should the claim that E_S confirms H_L.

White (2003, 244) makes a related point about the dubious metaphysical story we would need to support the parallel proposal about the FTA case. He considers two ways we could defend the claim that there is a probabilistic link between the existence of many (non-α) universes and my own existence. First, we might "suppose that I was once an unconscious soul waiting to be embodied in whichever universe produced a hospitable living organism." Alternatively, we might reason as follows: "The more universes there are, the more living creatures there are. So the more opportunities I had to be picked out of the pool of 'possible beings,' and hence the greater the likelihood that *I* should be observing anything." Such stories have no more plausibility in the FTA case than do the parallel proposals in the case of Fish.

A related way to try to defend the claim that there is a probabilistic link between H_L and Asha's presence in the lake turns on denying Kripke's thesis of the *necessity of origins*. I suggested that the existence of more large fish in the lake makes Asha's presence no more likely because she could only have been born to her actual parents. But this claim might be rejected: if Asha's existence is not tied to her specific parents, then (one might suggest) the presence of additional large fish does make it more likely that she will be born (and hence, be in the lake to be caught).

But rejecting the necessity of origins does not seem to make it any more plausible that there is a probabilistic link between the number of large fish and Asha's existence. If we take "being Asha" to be essentially a matter of having some exact genetic property, then we can suppose that it is not metaphysically necessary that Asha be born to her actual parents. Still, the chance that a fish with that exact genetic property would be in the lake seems unconnected to the total number of large fish: even if it is *metaphysically* possible for Asha to be born to different parents, a fish with that exact genetic property is simply not going to be born to different parents in any relevantly close possible world, such that the probability of Asha's being in the lake increases with the number of total large fish.

We need not insist that there is a *metaphysical* barrier to Asha's being born to different parents; the merely *empirical* fact that large fish outside Asha's direct lineage have no chance of giving birth to Asha means that H_L makes it no more likely that Asha would be in the lake. (The parallel point in the FTA case is that, even if we accept that it is metaphysically possible that we could have existed in a different universe, there is simply no *empirically* plausible hypothesis according to which *we* (the specific observers we are) would have come to exist in a different universe.)

The upshot is that we cannot salvage the intuitive verdict that we get confirmation in the case of Fish, so long as we insist on using the more specific version of our evidence (E_S). We really must reject RTE. And, when we turn to the second reference class problem—the question of how to characterize the *observer*—we should likewise reject the claim that we must always use the most specific reference class available. Here, the narrowest reference class is one that classifies the evidence as an observation of *mine*; and there is some initial plausibility to the idea that this reference class will always be the appropriate one to use in characterizing my evidence. The reason is that the narrowest reference class of observers is a reference class that is, in all *typical* cases, always salient antecedent to observing an outcome: when conducting experiments, or simply observing the world around us, we typically have a firm, antecedent grip on our own, unique identities. But in the case of FTA, it is precisely this typically-secure grip on our identities that is at issue, because the probabilistic process we are assessing—the world-creation event—is one on whose specific outcome our very identities as observers depend. Thus, the Predesignation Requirement that compels us to use the more general statement of the content of our observation in a simple case like Fish also compels us to use the more general characterization of our own identities as observers in the case of FTA.

Putting together our verdicts about how to characterize both the contents of our observations and the identity of the observer, we get the result that, properly applied, LP shows that we do indeed get confirmation of H_M over H_U from the evidence we have.

References

Bostrom, Nick. 2002. *Anthropic Bias: Observation Selection Effects in Science and Philosophy*. London: Routledge.

Bradley, Darren. 2009. "Multiple Universes and Observation Selection Effects." *American Philosophical Quarterly* 46:61-72.

———. 2012. "Four Problems About Self-Locating Belief." *Philosophical Review* 121:149-77.

Draper, Kai, Paul Draper, and Joel Pust. 2007. "Probabilistic Arguments for Multiple Universes." *Pacific Philosophical Quarterly* 88:288-307.

Eddington, Arthur. 1939. *The Philosophy of Physical Science*. Cambridge: Cambridge University Press.

Hacking, Ian. 1965. *Logic of Statistical Inference*. Cambridge: Cambridge University Press.

———. 1987. "The Inverse Gambler's Fallacy: the Argument from Design. The Anthropic Principle Applied to Wheeler Universes." *Mind* 96:331-40.

Hájek, Alan. 2007. "The Reference Class Problem is Your Problem Too." *Synthese* 156:563-85.

Juhl, Cory. 2005. "Fine-Tuning, Many Worlds, and the 'Inverse Gambler's Fallacy.'" *Nous* 39:337-47.

Leslie, John. 1988. "No Inverse Gambler's Fallacy in Cosmology." *Mind* 97:269-72.

Kotzen, Matthew. 2012. "Selection Biases in Likelihood Arguments." *British Journal for the Philosophy of Science* 63:825-39.

Kripke, Saul A. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Sober, Elliott. 2009. "Absence of Evidence and Evidence of Absence: Evidential Transitivity in Connection with Fossils, Fishing, Fine-Tuning, and Firing Squads." *Philosophical Studies* 143: 63-90.

Staley, Kent W. 2002. "What Experiment Did We Just Do? Counterfactual Error Statistics and Uncertainties About the Reference Class." *Philosophy of Science* 69: 279-99.

White, Roger. 2000. "Fine-Tuning and Multiple Universes." *Nous* 34: 260-76.

———. 2003. "Postscript to Fine-Tuning and Multiple Universes." In *God and Design: The Teleological Argument and Modern Science*, ed. Neil Manson, 243-50. London: Routledge.