Multiple Universes and Self-Locating Evidence

Yoaav Isaacs, John Hawthorne, and Jeffrey Sanford Russell *

Penultimate draft, June 2021

Abstract

Is the fact that our universe contains fine-tuned life evidence that we live in a multiverse? Hacking (1987) and White (2000) influentially argue that it is not. We approach this question through a systematic framework for self-locating epistemology. As it turns out, leading approaches to self-locating evidence agree that the fact that our own universe contains fine-tuned life indeed confirms the existence of a multiverse (at least in a suitably idealized setting). This convergence is no accident: we present two theorems showing that in this setting, *any* updating rule that satisfies a few reasonable conditions will have the same feature. The conclusion that fine-tuned life provides evidence for a multiverse is hard to escape.

1 A Question of Size

Reasoning about the size of physical reality is epistemologically fraught. This paper will explore what such reasoning involves.

Modern discoveries in cosmology are often taken to give powerful evidence that physical reality is a lot bigger than we would otherwise have supposed. Physics seems to be staggeringly inhospitable for life. Given a universe with our kind of laws, it would be extremely surprising for it to

*Thanks to Cian Dorr, Robert Hirsch, Benjamin Levinstein, participants in an epistemology seminar at the University of Southern California, and three anonymous referees for very helpful feedback. Thanks to Robert Hirsch for creating the diagrams. This publication was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

support life—and yet here we are.[1] It's commonly thought that this discovery is strong evidence that there are vastly many universes—a multiverse.[2] In the right kind of multiverse, even if each universe is overwhelmingly likely to be devoid of life, it is probable that there is life somewhere or other.[3]

Some philosophers contend that—far from being powerful evidence—these cosmological considerations are no evidence for the multiverse at all. Ian Hacking (1987) and Roger White (2000) have each argued that the reasoning above relies on a fallacy. Hacking offers this analogy:

> The kibitzer asks, 'Is this the first roll of the dice, do you think, or have we made many a one earlier tonight?' The gambler . . . says, 'Can I wait until I see how this roll comes out, before I lay my bet with you on the number of past plays made tonight?' The kibitzer, no fool, agrees, although charging a slight fee for allowing this extra 'information'. The roll is double-six. The gambler foolishly says, 'Ha, that makes a difference—I think there have been quite a few rolls.'
>
> [This gambler reasons] fallaciously. . . .
>
> The point is that the information available to the gambler is that double-six occurred at this throw. It is no more probable that double-six should occur at this throw, on the supposition of many previous throws, than it is that it should occur at this throw, on the supposition that this is the first throw tonight . . . (Hacking 1987, pp. 333–4).

Hacking contends that just as the occurrence of an improbable double-six on this throw does not make it make it more probable that there are many throws overall, the occurrence of improbably life-permitting conditions in *this* universe does not make it more probable that there is a multiverse. White (2000) develops and defends this line of argument (as we discuss in sections 2 and 6).

[1]See Weinberg (1989) for a technical analysis of issues relating to the cosmological constant. For a more generally accessible overview, see Lewis and Barnes (2017).

[2]For an overview of multiverse physics see Vilenkin 2011; Guth 2007.

[3]We are interpreting fine-tuning as telling us something explicitly probabilistic: see sections 5 and 7. There is some controversy about how one should read off probabilistic claims from facts about the values of the physical constants in parameter space (see inter alia McGrew, McGrew, and Vestrup 2001). But physicists are comfortable making such claims, and while this is not the place to evaluate the basis of those claims, for the purposes of this essay we are comfortable following them. See (Weinberg 1989, 2000) for discussion of the physics and (Hawthorne and Isaacs 2018) for philosophical discussion.

Objectors to Hacking and White offer competing analogies,[4] while their defenders marshal yet more analogies in reply.[5] But, as this flurry of conflicting analogies shows, we don't yet have a sufficient understanding of this issue to see which analogies are apt and which are not. More systematic theorizing is needed.

Fortunately, the last twenty years of epistemology are rich in resources that we can apply. The key issue is *self-locating evidence*, evidence concerning our own place in the world. If all we knew was the general fact that there is fine-tuned life, confirmation of a multiverse would be straightforward. Indeed, as we shall explain in section 3, nearly any mundane general facts support the multiverse hypothesis—and very specific general qualitative evidence, of the sort that we routinely learn through ordinary experiences, supports the multiverse hypothesis very powerfully. But Hacking and White call our attention to other evidence we have besides general qualitative facts: evidence not just about the existence of a certain sort of universe, but about our own universe, and about ourselves.

There is no consensus about how updating on such self-locating evidence should work; but there are some theories we can apply. In section 4 we introduce three leading approaches, and in section 5 we show that all of them agree on the central question: the fact that our own universe contains fine-tuned life does indeed confirm the existence of a multiverse. (However, only two of the three theories say that *fine-tuning* makes any difference to the strength of this evidence, beyond the simple fact that we are alive at all: see section 6.) It is no accident that the leading theories agree on the central question. In section 7 we present two theorems that show that *any* rule for generating posterior probabilities that satisfies certain reasonable constraints (and some simplifications and idealizations) will lead to the same conclusion. The conclusion that fine-tuned life provides evidence for a multiverse is hard to escape.

This essay is quite long, and different readers will be most interested in different parts. Readers who are interested in our dialectic with Hacking and White will want to read sections 2 and 6, where we give our replies to their arguments. Section 4 is aimed at catching readers up who are rel-

[4]For some examples: other dice rolls (McGrath 1988), a photographer who visits casinos (Whitaker 1988), firing squads (Leslie 1988, 1989), school prizes (Holder 2002), and more dice (Bradley 2009).

[5]To yet more dice (Rota 2005), a kidnapper with a mysterious card-sorting device (Juhl 2006), the decay of named uranium atoms (Draper, Draper, and Pust 2007), where in Brooklyn one's parents lived (Leeds 2007), trying on off-the-rack suits (Landsman 2016) (this is a turned-around version of one from (Rees 2008)), and more.

atively unfamiliar with views on self-locating evidence. Section 5 applies these theories to fine-tuning and the multiverse; for those who are already comfortable with the theories in question (Compartmentalized Conditionalization, Self Indication, and Self Sampling) the summary in table 1 at the end of section 5 may suffice. Some readers may wish to go directly to our most general results (which subsume those in section 5), and then read back as needed: the main theorems can be found in section 7 and appendices A and B.

2 This Universe

Hacking and White each argue that our evidence concerning fine-tuned life does not support a multiverse. Let's start by saying what is right about their criticism.

As all parties concede, the general qualitative proposition *there is fine-tuned life* is straightforwardly evidence for the multiverse. (We discuss this in more detail in section 3.) But this cannot settle the issue by itself. Our relevant evidence is not just the general evidence that there is fine-tuned life in some universe or other: rather, we know the more specific fact that there is fine-tuned life in *this* universe, *our* universe. As White points out, it is illegitimate to ignore the specific evidence and only pay attention to the weaker general evidence. Specific evidence can *screen off* the import of general evidence.

> . . . [O]f course the more universes there are, the more likely it is that *some* universe supports life. That is, M [there is a multiverse] raises the probability of E' [there is fine-tuned life] . . . . But now, the response goes, we know that E' is true since it follows from E [α contains fine-tuned life, where α is our own universe]. So E' confirms M even if E does not. In other words, our knowledge that some universe is life-permitting seems to give us reason to accept the Multiple Universe hypothesis, even if our knowledge that α is life-permitting does not.
>
> We can quickly see that there is something going wrong here. . . . Suppose I'm wondering why I feel sick today, and someone suggests that perhaps Adam got drunk last night. . . . Perhaps if all I knew (by word of mouth, say) was that someone or other was sick, this would provide some evidence that Adam got drunk. But not when I know specifically that I feel sick. This suggests

> that in the confirming of hypotheses, we cannot, as a general rule, set aside a specific piece of evidence in favor of a weaker piece. (White 2000, p. 264)

All of this is correct.

But Hacking and White each go further, and make a positive claim about the import of our specific evidence ("this universe contains fine-tuned life"): namely, that it is independent of how many universes there are, and does not confirm the hypothesis that there are many universes rather than one. This conclusion, we contend, is premature. We just don't understand how singular evidence like this works nearly well enough be confident in such judgments. What's more, we will argue (in sections 5 to 7) that there are strong systematic reasons to think that this further claim is false. We will make the case that that our specific evidence *does* confirm a multiverse.

There are many difficult issues about the probabilities of singular propositions. White's main argument illustrates some of these difficulties (2000, 262–263). White writes, "Let α be our universe and let T_1 be the configuration which is necessary to permit life to evolve." He then considers the claim "α instantiates T_1":

> . . . [T]he probability of this is just $1/n$, regardless of how many other universes there are, since α's initial conditions and constants are selected randomly from a set of n equally probable alternatives, a selection which is independent of the existence of other universes.

But this conclusion does not follow, for subtle reasons. One of the dangers of *de re* probability ascriptions is that they are susceptible to Frege puzzles (see for example Chalmers 2011). We can grant what White insists on:

> The name 'α' is to be understood here as *rigidly designating* the universe which happens to be ours. Of course, in one sense, a universe can't be *ours* unless it is life-permitting. But the universe which happens actually to be ours, namely α, might not have been ours, or anyone's. It had a slim chance of containing life at all. (White 2000, p. 274, note 6, original emphasis)

But this still does not settle the issue.

Consider a simple toy model. There is either one universe stamped with the label *Universe One*, or two universes stamped with the labels *Universe One* and *Universe Two*, respectively. (The labels are stamped on the outside, so people inside a universe can't see them.) If there is one universe, then the objective chance that Universe One has life is $1/n$. If there are two universes, then the chance that Universe One has life is $1/n$, and the chance that Universe Two has life is $1/n$, and these are independent. All of these facts are known. Roger is in Universe One, but for all he can tell he might be in Universe Two. Roger says "Let α be my universe." So "α" rigidly designates Universe One—though Roger does not know this.

The *objective chance* that α would contain life is $1/n$. But it does not follow in this case that the *prior epistemic probability* of "α contains life" is $1/n$. We must be extremely careful when we try to apply chance-credence principles to de re probabilities: *chance* statements are referentially transparent; *epistemic probability* statements are not (see Hawthorne 2002; Hawthorne and Lasonen-Aarnio 2009).[6] For example, one might introduce a name like this: "Let β be Universe One if Universe One contains life or there is just one universe, and otherwise Universe Two." In this case, β in fact *is* Universe One, and has chance $1/n$ of containing life. But the prior epistemic probability of "β contains life" is the same as the objective chance that *either Universe One or Universe Two contains life*—which is clearly higher than $1/n$.

The prior probability of "Universe One contains life" is $1/n$. But the prior probability of "α is Universe One" is less than one (even though this is a metaphysically necessary truth): Roger can tell that he is in α, but he can't tell that he is in Universe One. So it does not follow from the fact that "Universe One has life" has prior probability $1/n$ that "α has life" has prior probability $1/n$. The correct prior probability is difficult to determine. In particular, it is obscure whether this is independent of how many universes there are. (There are two reasons "α" is much less clear-cut than "β". First, the way the name "α" was introduced leaves it unclear what it refers to, if anything, in cases where Roger is not in any universe—for example, cases in which no universe contains any life at all. Second, it is unclear what prior probability Roger should assign to being in one universe or another. We discuss the second issue in section 6 and appendix B.)

In a later postscript, White (2003, p. 244) writes about his use of the proper name α,

> I admit that there are difficult issues here in which I would rather not get entangled, and I regret putting the argument in these terms as I now think the crucial issue is independent of

[6]In our discussion of de re probabilities in this section we will speak as if *sentences* are the bearers of probability—since if α is Universe One, the *proposition* that α contains life may well be the same as the proposition that Universe One contains life. In section 4 we will introduce a different way of avoiding this difficulty for our more official framework.

these matters . . .

He then dispenses with the proper name and proceeds in *first-personal* terms.[7] This is a good idea. Instead of the evidence one might state with "α has fine-tuned life" (where "α" has been introduced as a proper name for this universe) or "This universe has fine-tuned life," we will focus on evidence one might state with "*I* am in a universe that contains fine-tuned life." (Each of us has evidence we can express with these words; we do not presuppose that it is the *same* evidence in each case. We handle this more carefully in section 4.) It would be surprising if evidence stated in terms of "*this* universe" and evidence stated in terms of "*my* universe" had dramatically different import. The standard label for evidence about oneself is *self-locating* evidence.[8]

This narrows the scope of the problem, but it does not make it easy. The difficulties of probabilistic self-locating epistemology are well-known (for overview, see Meacham 2008; Titelbaum 2016; Manley, Manuscript). There is nothing like a consensus about how this works. But we are not utterly at a loss: we have some proposals we can apply, and a framework in which we can make the questions precise.

It is worth clarifying what White means, and what we mean, by the question of whether something *confirms* the multiverse. The standard Bayesian picture is that evidential support means probability-raising—but with respect to what? The picture that has been assumed is that there are certain *prior epistemic probabilities*, and that to confirm a hypothesis (in the sense at issue) is to raise its probability above its prior probability. Note that these "priors" cannot be literally understood as the credences that you had at some earlier time: for the priors are unopinionated about whether there is life, or concrete agents, or any complex matter. It is difficult to make sense of an agent (even a highly idealized agent) who is so deprived of evidence as to be unopinionated about such questions as these. Instead, we are thinking about the priors as epistemic *ur-priors*.[9] Such probabilities encode rela-

[7]"If my observation is to provide *me* with evidence of these other rolls, they will have to make it more likely that *I* would observe this" White (2003, p. 245, original emphasis). We take up his ensuing argument in section 6.

[8]Some philosophers pursue the ambitious project of reducing *all* attitudes toward singular propositions to self-locating attitudes (see Lewis 1979; Chalmers 2011; Ninan 2013; for critical discussion see Cappelen and Dever 2013; Magidor 2015; Yli-Vakkuri and Hawthorne 2018). We need not take any stand here on this ambitious project's prospects. Stalnaker (2008) and Moss (2012) use the same kind of trick in the other direction, reducing questions of self-locating probabilities to questions about probabilities of propositions stated using special names or demonstratives.

[9]This idea arises from (Keynes 1921) and (Carnap 1950) and is defended by (Williamson

tions of evidential support between qualitative propositions; it is natural to model these using objective physical chances, imagining, as White said, that the universe's "initial conditions and constants are selected randomly." In order to sidestep the complications we just discussed concerning *de re* probabilities, for most of this paper we will only suppose there to be ur-prior probabilities for general qualitative propositions—such as the proposition that there are many universes, or that there is life. We do not assume that our posterior probabilities can be calculated simply by conditionalizing the ur-priors: rather, we will be considering a variety of updating rules for self-locating evidence, which somehow or other take you from *qualitative* priors and *self-locating* evidence to generate self-locating posterior probabilities.[10] It may be a vague matter what the exact ur-priors are; it may also be a "subjective" matter in some sense. We will not take up these foundational issues further here.

3 Qualitative Evidence and Multiverses

Before we consider theories of self-locating evidence, let's take a step back and consider the import of our qualitative evidence a bit more carefully. For the moment, we set self-location aside.

You know that you're reading a philosophical argument right now. That's a non-qualitative proposition. So for the time being we'll pretend that you only know something weaker and qualitative—that someone reads a philosophical argument at some point. Given this pretense, everything is technically tractable: qualitative hypotheses can straightforwardly be assigned prior probabilities, and updating on qualitative evidence can be done by straightforward conditionalization. But doing so yields alien results.

Let's make some simplifying assumptions. Suppose we know that either (1) there is a single universe, or (2) there is a multiverse containing 10^{100} universes. Suppose that our expectations about the intrinsic properties of any given universe are the same whether it's a solitary universe or part of a multiverse, and also that what goes on in one universe in a multiverse is probabilistically independent of what goes on in each other universe in that multiverse.[11] Suppose furthermore that each universe is finite, extending

2000); for overview see (Meacham 2016).

[10]In appendix B we also consider the more ambitious idea of *self-locating* ur-priors, and the relationship between these rules and conditionalization. (Arntzenius and Dorr 2017); see again (Meacham 2016).

[11]Compare our discussion of *separable priors* in section 7 and appendix A.

no more than a quadrillion (10^{15}) light years across and lasting for no more than a quadrillion years. Suppose that the prior probability of a multiverse is just one in a million.

What happens to that probability given our qualitative evidence? You might expect that the only relevant evidence will involve technical details from physics. But in fact mundane evidence looms even larger.

To take a famous example, Borges (1941) imagined a library containing all possible 410-page books of a particular format and alphabet. Such a library must be truly gargantuan, containing well over 10^{4000} books—staggeringly too many to fit in a single universe with the specified dimensions.[12] Suppose you read one of Borges's books[13]—one of the very many "random-looking" books, not a long string of the letter q, or the text of *Paradise Lost*. The qualitative evidence you gain is just that someone reads a 410-page book of such-and-such qualitative profile at some point. Given a solitary universe, this evidence is ludicrously unlikely: only a minuscule proportion of all Borges-books can be found in a single universe. Given a multiverse, this evidence is *still* ludicrously unlikely. There are so many Borges-books—many orders of magnitude more than 10^{100}—that only very few of them can be found even in a large multiverse. But the tiny probability of finding this Borges book somewhere in 10^{100} universes is nearly 10^{100} times greater than the even tinier probability given just one shot. So, given our assumptions, the fact that someone reads a book with some particular random-looking qualitative profile is staggeringly strong evidence for the multiverse. This evidence alone raises the probability of a multiverse from one in a million to about $99.9\cdots999\%$, where that's a string of 94 nines.

In fact, the qualitative evidence gained by reading nearly *any* book will overwhelmingly support the multiverse. A 410-page history of the Napoleonic wars will have a similar effect. Such a book is substantially more likely to be found in a single universe than 410 pages of specific gibberish, but still extremely unlikely: given reasonable assumptions, the probability that a universe would contain any book containing those particular words in that order is still much less than one in 10^{100}.[14] So it is still nearly 10^{100}

[12]If we assume that each book is at least the size of a postage stamp and lasts at least a nanosecond, and that no two books have overlapping spatiotemporal envelopes, then there are well under 10^{150} distinct books per universe.

[13]There's a charming website that generates them: `https://libraryofbabel.info/`

[14]Even if each page of the book only contains a hundred words, and each word only provides an average of one bit of new information, there are still more than 10^{500} 410-page histories of the Napoleonic wars in Borges' library. Again only a tiny proportion of these books fit in a single universe.

times more likely that someone would read that sequence of words in a multiverse of 10^{100} universes than in a single a universe. Similarly, listening to music, eating sandwiches, or paying attention to just about anything else in ordinary life is liable to provide qualitative information that is so specific that it is massively unlikely for things *just like that* to arise in a single universe. The only goings-on that will *not* massively confirm a multiverse are those which have a reasonable shot at going on in any single universe. And even such extremely banal goings-on still count in favor of the multiverse at least a little, unless they are certain to occur, come what may.

The root issue here is that our qualitative evidence is ordinarily about what there is rather than what there isn't. You can learn just by looking that there are brown horses. You can't ordinarily learn just by looking that there are no blue horses. Of course, you can learn that there are no blue horses *near you*, but that's not a general qualitative proposition.[15] If the universe is big and random enough, then it's nearly sure to contain blue horses. If you did somehow learn that there are—without restriction—no blue horses, that would be overwhelming evidence against the existence of a large and random multiverse. But it's hard to see how one would learn that sort of thing.

We can make this point more precise. For worlds w and w^+, let a *qualitative embedding* of w in w^+ be a one-to-one function from the concrete objects in w to concrete objects in w^+ which preserves all intrinsic properties and relations; in this case, say that w^+ *embeds* w. Intuitively, w^+ includes a qualitative copy of everything in w, and perhaps more besides. Call a proposition p *local* iff for any world w in which p is true, p is also true in any world w^+ that embeds w.[16] Similarly, let a *local property* be one that is preserved by embeddings of one world in another.

Any local proposition that is true in a single universe is also true in any multiverse that includes a copy of that single universe. Whatever probability such a proposition may have conditional on there being a single universe (strictly between zero and one), the probability must be greater conditional on there being many universes. (This still supposes that the intrinsic qualitative profiles of each universe are independent and identically

[15]One complication is that natural kind terms like "horse" are plausibly not really qualitative either. Let this discussion be officially understood as concerning qualitative horse-duplicates.

[16]For propositions expressed in a first-order language where all predicates stand for intrinsic properties or relations, local propositions are precisely those expressed by $\exists_1$ sentences: those that consist of a string of existential quantifiers in front of a quantifier-free formula. (See Russell 2020).

distributed.) It follows that *local* qualitative evidence can't help but support a multiverse—and the more specific the evidence is (that is, the smaller its prior probability), the more powerfully it supports a multiverse.

How can it be that our qualitative evidence can count strongly in favor of a multiverse, but can't count against it? This is due to an asymmetry in what this qualitative evidence can be like. Any non-trivial local qualitative proposition q counts in favor of a multiverse. In that case, not-q would count as evidence against a multiverse. But we have supposed that while q is the sort of proposition one can *have* as evidence, not-q is not: the negation of a local proposition is generally not a local proposition itself. *There are blue horses* is local, but its negation *there are no blue horses* is not.

This kind of powerful confirmation of extravagant hypotheses by mundane evidence is unsettling. Fine-tuning evidence works this way, too. (That is, the *qualitative* aspect of fine-tuning goes the same way. It remains to be seen whether fine-tuning makes a special difference to *self-locating* evidence.) It tells us that the existence of life and complexity is very unlikely given a single universe—much, much less likely than one might have naïvely expected. This gives us reason to think that some very general local facts about our universe—such as the existence of life, stars, or tungsten—have extremely low prior probability given a single universe. If the probability of life given a single universe is p (and each universe's chance at life is independent of the others), then the existence of life improves the odds of the multiverse hypothesis by a factor of $(1 - (1 - p)^n)/p$. For large n this approaches $1/p$. If p is very small, this is very powerful confirmation.

But even without fine-tuning, we already had *some* evidence with this same feature: local qualitative facts that have very low prior probability, given a single universe, simply due to including a lot of detail.

Can massive confirmation of a multiverse be escaped? In this section we've been ignoring self-locating evidence. This suggests that it may be surprisingly important that we take self-locating evidence into account! You don't just know that *someone* read such-and-such book, listened to such-and-such music, ate such-and-such a sandwich, and so on. You know that *you* read such-and-such book, listened to such-and-such music, ate such-and-such a sandwich, and so on. This self-locating evidence might help. It *might*—as Hacking and White claim—screen off the qualitative evidence from multiverse hypotheses. It's not obvious whether it does. But self-locating evidence is our most plausible hope for avoiding the overwhelming confirmation of extravagant hypotheses.

4 Three Rules for Self-Locating Evidence

Let us be very clear: there is no settled method for the epistemology of self-locating evidence. All of the precise theories we know of face very serious objections. There is no final science we can present here, but we will begin by introducing three leading approaches to the problem. In section 7 we present general results that encompass many alternative self-locating epistemologies besides these three.

Of course, one can easily find many more than three theories of self-locating belief in the philosophy literature. (For helpful overview see Titelbaum 2013, 2016; for a few more recent approaches see also Meacham 2016; Arntzenius and Dorr 2017; Schwarz 2017; Builes 2020). But many of the differences between these theories are in their answers to questions that don't immediately arise here: questions about the passage of time, forgetting information, or people or universes that undergo fission and branching. None of these tricky issues are immediately relevant here: while they *can* arise for agents thinking about fine-tuning, they don't have to. For the sake of clarity, we will begin by focusing on simple cases where they do not arise. In the end, some of the additional complications may turn out to be relevant. Branching universes—which feature in some interpretations of quantum mechanics—seem especially relevant for a complete account of multiverse epistemology (see inter alia Greaves 2004; Bradley 2011; Wilson 2013; Sebens and Carroll 2016). The general results we present in section 7 are less sensitive to the simplifications we will make here.

We are only considering propositions that can be expressed using sentences with just the words "I", "now", and qualitative vocabulary. We can think of any such sentence as having the canonical form "I am now *F*" where *F* expresses a qualitative property.[17] It is helpful to model things by focusing on the properties, rather than the propositions. If you have the evidence you would express with the sentence "I am now happy," we can say your self-locating evidence includes the property *being happy*. While we do not wish to presuppose that we each express the same *proposition* with this sentence whenever we say it, many people can self-ascribe this *property* on many occasions. We can handle ordinary qualitative evidence as a special case: we can associate each qualitative proposition p with the "boring" qualitative property *being such that p*.[18]

[17] Again, we do not wish to presuppose that *all* propositions can be reduced to propositions expressed this way.

[18] In general, we should not presuppose that *being such that p* requires *being someone*; we discuss this in appendix B, which is the only place it makes a difference.

As it is convenient to represent propositions with sets of worlds, it is convenient to represent properties of agents with sets of *centers*, which are triples $\langle a, t, w \rangle$ of an agent, a time, and an epistemically possible world. (We will think of *worlds* as specified in qualitative and eternal terms.) For example, *being happy* is represented by the set of triples $\langle a, t, w \rangle$ of an agent a who is happy at time t in world w. So one's total self-locating (and qualitative) evidence can be represented as a set of centers: if one's total self-locating evidence is that one is F, then these are the centers that represent agents who are F.[19] We will call these *live* centers, and centers that are not in this set *dead*.

To keep things tractable, we will suppose there are only finitely many distinct worlds, each of which contains finitely many centers.[20] We start with a *prior* probability distribution over these worlds; these priors assign probabilities to qualitative matters, but are silent on self-locating matters. (As we discussed in section 2, we are thinking of these as cosmological "ur-priors" rather than actual credences someone had at an earlier time.) We will consider various rules that take a qualitative prior, together with self-locating evidence (represented by a set of live centers), and produce *posterior* probabilities for each world. The posterior probability of any qualitative proposition can then be calculated by adding up the probabilities of the worlds in it. The rules work by assigning posterior probabilities to each *center*, and thereby to possessing any particular qualitative property (by adding up the centers in that property). But our main focus will be on the probabilities of qualitative hypotheses—like the hypothesis that there are many universes.

Consider a really simple situation. There are three worlds. World 1 has someone in a black room (and no one else). World 2 has someone in a white room (and no one else). And world 3 has two people in a black room and one person in a white room (and no one else). Let's suppose that each world has prior probability 1/3. You find yourself in a black room: *being in a black room* is your total self-locating evidence. What posterior probabilities should you assign to each world?

Here's the first method for assigning probabilities.[21]

[19]This builds in the implicit assumption that one's evidence includes being an agent.

[20]This assumption is relaxed somewhat in appendices A and B. Worlds containing infinitely many agents are of serious interest in modern cosmology. An important open question is the so-called "measure problem," which in effect amounts to the problem of finding an appropriate self-locating epistemology for infinite worlds (for overview see Vilenkin 2011; Guth 2007, sec. 4).

[21]Halpern 2004; Meacham 2008; see also Builes 2020.

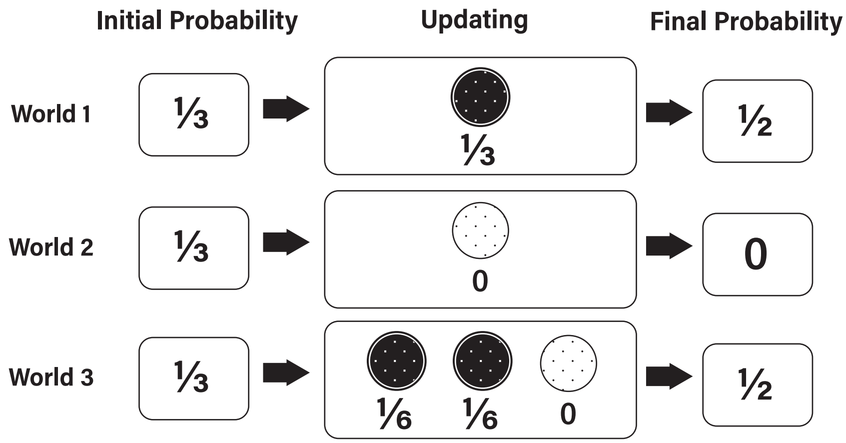

Figure 1: Compartmentalized Conditionalization.

Compartmentalized Conditionalization.

1. For each world w that contains any live centers, divide the probability of w evenly between them: that is, assign each live center probability p/n, where p is the prior probability of w and n is the number of live centers in w.
2. Assign each dead center probability zero.
3. Renormalize so everything adds up to one.

When it comes to qualitative propositions, Compartmentalized Conditionalization for the evidence *being F* amounts to standard conditionalization on the qualitative proposition *someone is F*.

Let's work through the simple case (fig. 1). The world containing no one in a black room is ruled out, and the probabilities of the remaining worlds are renormalized. So the agent winds up having credence 1/2 in world 1 and credence 1/2 in world 3.

The main motivating feature of Compartmentalized Conditionalization is the idea that the probability of qualitative hypotheses is unaffected by purely self-locating information: the support that being F gives to a qualitative hypothesis is precisely the same as that of the qualitative proposition *someone is F*. This same feature, though, makes Compartmentalized Conditionalization hopeless for avoiding confirmation of multiverses. As we discussed, non-trivial propositions of the form *someone is F* (where F is any local property) will always confirm large random multiverses over single universes. According to Compartmentalized Conditionalization, self-locating evidence does nothing for qualitative hypotheses beyond what these qualitative propositions do.

There are serious objections to Compartmentalized Conditionalization besides this. There are other reasons to think that self-locating information *should* make a difference to the probability of qualitative hypotheses. Consider a simple example adapted from Ruth Weintraub (Weintraub 2004; see also Bostrom 2002; Titelbaum 2008; Briggs 2010; Dorr, Manuscript).[22] There will certainly be three people. A fair coin is flipped. If the coin is heads, one person sees a red light, and two see a green light. If the coin lands tails, two people see a red light, and one sees a green light. It seems clear that seeing a red light is evidence that the coin came up tails. (Indeed, it seems clear how strong this evidence is: one is twice as likely to see a red light if the coin came up tails, so the posterior odds should be 2 : 1.) But the proposition *at least one person sees a red light* is evidence for neither hypothesis.[23] Compartmentalized Conditionalization has few defenders in either physics or philosophy.

Michael Titelbaum's (2008; 2012; see also 2016, p. 674) "Certainty Loss Framework" (CLF) says that Compartmentalized Conditionalization is correct in certain restricted cases: in our terminology, these are cases where, whatever one's evidence might be, there is at most one live center in each possible world.[24] This restriction avoids the apparent counterexamples of the kind we just considered. But this restricted rule falls silent on the mul-

[22]Weintraub's original case involved a single person at three times, with memory loss after each flash of light.

[23]Sometimes Compartmentalized Conditionalization also recommends *changing* one's view in odd ways. Consider this example. Initially, with objective chance 1/2 there is one person in a black room, and with chance 1/2 there are two people in separate black rooms. Each person flips a fair coin and observes how it lands. According to Compartmentalized Conditionalization, *before* the coin flip one should have credence 1/2 in being the only person, but *after* the coin flip (however it lands) one should assign credence 2/5 to being the only person.

[24]"CLF allows an agent to update by Compartmentalized Conditionalization only if for each time and each centered proposition she entertains, the agent has some uncentered proposition she is certain at that time has the same truth-value as the centered proposition" (2016, p. 674). One complication for applying Titelbaum's theory in our setting—or any theory that is engineered primarily as a *diachronic* constraint on agents—is that the "priors" we are considering over cosmological hypotheses are not literally the prior credences of any agent. Here we are extending Titelbaum's theory to apply to these ideal "ur-priors" so that we can derive lessons about fine-tuning; but Titelbaum himself may not welcome this extension.

tiverse cases we are interested in, which can involve multiple agents in the same world with the same self-locating evidence. Titelbaum's strategy for deriving verdicts for cases that do not satisfy the single-center constraint is to find analogous cases that do satisfy it. We think that this strategy can be applied to multiverse models we consider in section 5; when carried through, the upshot is that Titelbaum's CLF approach delivers the very same verdicts in these cases as the next rule we will consider: *Self-Indication* (see footnote 29). (In cases where the single-center restriction *does* hold, Compartmentalized Conditionalization and Self-Indication give precisely the same results.) So we will not consider CLF as a separate rule in what follows.

Here's the second method for assigning probabilities.[25]

Self Indication.

1. Assign each center the probability of the world it is in.
2. Reassign each dead center probability zero.
3. Renormalize so everything adds up to one.

When it comes to qualitative hypotheses, the effect of Self Indication is that the relative probability of each world w is boosted by the number of live centers in w. That is, we can calculate the posterior probability of w by multiplying $\Pr(w)$ by the number of live centers in w, and then renormalizing.

Let's work through the simple example again (fig. 2). First, each of the centers representing an agent in a black room is assigned 1/3 (the probability of its world). Second, the centers representing agents in white rooms are zeroed out. Finally, we renormalize. So the agent winds up with credence 1/3 in world 1 and credence 2/3 in world 3. Since there is one live center in world 1, zero live centers in world 2, and two live centers in world 2, we eliminate world 2 and boost the relative probability of world 3 by a factor of two (from 1 : 1 odds to 2 : 1 odds).

Unfortunately, Self Indication does not escape the result that mundane evidence overwhelmingly confirms multiverse hypotheses, either. Self Indication confirms a hypothesis in proportion to its expected number of live centers—the more live centers, the more confirmation. For any local

[25]Bostrom (2002, p. 66) uses the name the "Self-Indication Assumption" for this principle: "Given the fact that you exist, you should (other things equal) favor hypotheses according to which many observers exist over hypotheses on which few observers exist." For versions of this idea promoted by physicists, see Vilenkin (1995, eq. (1) on p. 847) and Olum (2002).

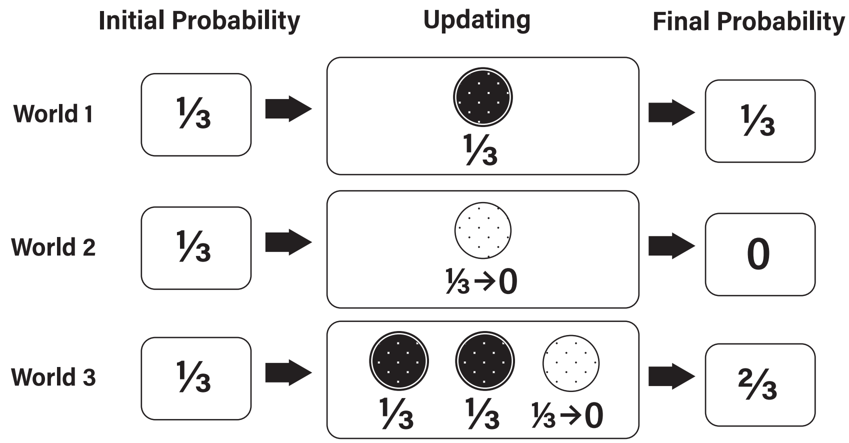

Figure 2: Self Indication

property F, if the expected number of such centers in a single universe is n, then the expected number of such centers in a multiverse of a trillion independent universes is a trillion times n—so this kind of multiverse is confirmed by a factor of a trillion over a single universe. This does not solve the problem of mundane evidence confirming extravagant hypotheses.

We'd like to note that, while it is counterintuitive, there is a certain sense to this kind of confirmation of extravagant hypotheses. Putting things very casually, one might think that the more people there are, the more likely it would be that you are one of them. It is thus fairly natural to think that your existence is evidence for there being more people. We're not saying that we're happy with this verdict. But we are saying that unhappiness with it shouldn't be considered decisive.

Here's the third method for assigning probabilities.[26]

Self Sampling.

1. For each world w that contains any centers at all, divide the

[26]Bostrom (2002, p. 57) uses the label "Self Sampling" for this principle: "All other things equal, an observer should reason as if they are randomly selected from the set of all possible observers." Note that this original use of the term does not pick out a specific rule; we use the term in a more specific sense, combining Bostrom's idea with the idea that there is no additional re-weighting of qualitative worlds. For Self Sampling reasoning in physics, see Page (1999, especially pp. 226–227) (However, this example is complicated by considerations about quantum measures, which we return to in section 7).

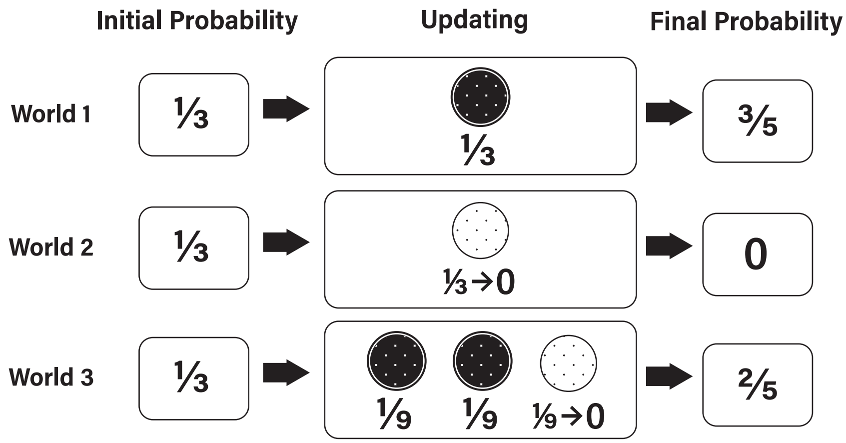

Figure 3: Self Sampling

probability of w evenly between all of them: that is, assign each center probability p/n, where p is the prior probability of w and n is the total number of centers in w.

2. Reassign each dead center probability zero.
3. Renormalize so everything adds up to one.

The effect is that the relative probability of each world w containing any centers is boosted according to the *proportion* of live centers among all centers in w.

Let's work through the simple example one more time (section 4). First we assign each of the centers in world 1 and world 2 probability 1/3, and we split the probability of world 3 between its three centers, giving each 1/9. Then we reset the white-room centers to zero, and renormalize. So the agent winds up having credence 3/5 in world 1 and credence 2/5 in world 3.

Unlike the other two rules, Self Sampling does *not* have the effect that mundane local evidence confirms vast worlds. Self Sampling has a compensating mechanism, already on display in the simple example: if a world contains centers that *don't* fit your evidence, then these take away from the probability of the centers that *do* match your evidence. Suppose that each universe has the same probabilities of producing centers of any particular kind, regardless of what other universes there may be, and suppose you have the evidence *I am F*, where *F* expresses a specific enough local property so that it is unlikely that anyone is *F* in a single universe. If the issue is merely that *F* is specific, then it can still be very likely that there are agents in a single universe who are other ways *besides F*. (That is, we are not here considering a fine-tuning scenario in which it is unlikely that there is anyone at all in a single universe.) In that case, the multiverse hypothesis will make it much more probable that someone is *F*—but one should also expect, given a multiverse, that there will be vastly more agents who are *not F*. The Self Sampling rule weighs each world containing agents according to the proportion of *F* agents among all agents in that world. There is no general reason why the expected proportion among multiverse worlds containing agents should be any different from the expected proportion for a single universe containing agents. So unlike the other rules we have considered, Self Sampling need not confirm multiverses given just mundane local evidence.

But Self Sampling has other problems. One objection is the famous "Doomsday argument" (Carter 1983; Leslie 1990; see also Bostrom 2002, pp. 89ff. and references therein). Suppose that Earth is the only populated planet, and that there is a doomsday device that will destroy all life on earth in the year 2200 with a known objective chance of one in a million. Suppose that if the human race is not destroyed, we are guaranteed to endure for a billion years, stretching out beyond our galaxy and colonizing the universe. The odds seem good. But Self Sampling tells us to be very confident that the doomsday device will go off. Suppose that whether or not the doomsday device goes off, you can tell who you are and when it is: your evidence rules out you being in any other person's predicament, or now being any other time. So in each doomsday world, and likewise in each glorious-future world, there is just one center compatible with your self-locating evidence. But in a world with more centers total, the single *live* center receives a much smaller share of that world's probability. Since there are more centers in glorious-future worlds than doomsday worlds by a factor of billions, the posterior probability of the doomsday hypothesis is boosted from one in a million to close to one.

A second problem for Self-Sampling is that it is not entirely clear what it says: its recommendations are sensitive to what exactly counts as a *center* (in a way that those of the other two rules are not). Should we just count humans, or also pangolins or ants? What about ant colonies, or supercomputers, or proper parts (or temporal parts) of people? This is the so-called "reference class problem" (see Bostrom 2002, chs. 10–11; see also

Arntzenius and Dorr, Manuscript).[27]

We can summarize all of these rules as follows. Each rule has the effect of multiplying the prior probability of each world w by some "confirmation factor" $\eta(w)$, and then renormalizing. The confirmation factors for each update rule are as follows.

Compartmentalized Conditionalization

$$\eta_{CC}(w) = \begin{cases} 1 & \text{if there are any live centers in } w \\ 0 & \text{otherwise} \end{cases}$$

Self Indication

$$\eta_{SI}(w) = \text{the number of live centers in } w$$

Self Sampling For worlds w that contain any centers,

$$\eta_{SS}(w) = \text{the proportion of live centers in } w \text{ among all centers in } w$$

Otherwise,

$$\eta_{SS}(w) = 0$$

That tells us everything we need to know about the posterior probabilities of qualitative propositions. When it comes to self-locating probabilities, each of these rules simply says to evenly divide the posterior probability of each world w among the live centers in w.

We can sum that up even more simply. Consider a set of mutually exclusive qualitative hypotheses. Compartmentalized Conditionalization relatively confirms each hypothesis H in proportion to the conditional probability that there are *any* live centers, given H. Self Indication relatively confirms each hypothesis H in proportion to the *expected number* of live centers conditional on H. And Self Sampling relatively confirms each hypothesis H according to the *expected proportion* of live centers among all centers, where this proportion is considered to be zero in worlds containing no centers at all.

[27]The reason this problem does not arise for other views is that the number of *dead* centers makes no difference to the final probability. So, for example, if your evidence includes that you are human, pangolins and supercomputers don't matter to the final calculation.

5 Fine-Tuning and Multiverses

We now have three rules for dealing with self-locating evidence. Let's see what each of them tells us about the significance of fine-tuning for the multiverse. (Again, we will generalize the results of this section in section 7.)

We'll start by working through some simple toy models. Suppose that the prior probability that there is exactly one universe is 2/3, and the prior probability that there is a multiverse consisting of exactly four universes is 1/3. Suppose that each "universe" consists of a single room, containing at most one agent (at just one time). Each room is one of four colors—black, white, red, or green—each with equal prior probability. As in section 3, we suppose that what each universe is like—specifically, its color—is independent of what any other universes are like, on the prior. Each agent can tell what color room they are in, and nothing more. You find yourself in a black room. This stands in for your "mundane" (but reasonably detailed) evidence.

First, let's get a feel for how things work out if life doesn't need anything like fine-tuning: suppose that every universe is guaranteed to contain one agent. We'll add in fine-tuning after that. (The results of these calculations are summarized in table 1 on page 24.)[28]

Compartmentalized Conditionalization. The live centers correspond to the agents in black rooms. Given a single universe, the probability of there being such an agent is 1/4. Given a multiverse, the probability of there being at least one such agent is $1 - (3/4)^4 \approx 0.68$. So the multiverse is relatively confirmed by a factor of ≈ 2.7, bringing the probability from 1/3 to about 0.57. (From 1 : 2 odds to $\approx$ 2.7 : 2 odds.)

Self Indication. Each hypothesis is relatively confirmed in proportion to its expected number of agents in black rooms. Given a single universe, this expected number is 1/4. Given a multiverse, this is $4 \cdot (1/4) = 1$. So the multiverse is relatively confirmed by a factor of 4, bringing its probability from 1/3 to 2/3. (From 1 : 2 odds to 4 : 2 odds.)

Self Sampling. Since every world contains at least one agent in this model, each hypothesis is weighted according to its expected proportion of

[28]There are various measures of strength of confirmation (see Fitelson 1999). Here we focus on the Bayes factor, which is the factor by which the odds of a hypothesis is increased by evidence. (This is the same as the *log-likelihood ratio* Fitelson discusses, except on a linear rather than a logarithmic scale.) Nothing important turns on this choice.

black-room agents among all agents. Whether there is a single universe or a multiverse, this expected proportion is 1/4. So the multiverse hypothesis is not confirmed, and its posterior probability stays at 1/3.

That was straightforward enough. Now let's get a feel for how things work out if life is fine-tuned. Again, we'll use a simple toy model. As before, we let the prior probability of a single universe be 2/3, and the probability of a four-universe multiverse 1/3. As before, the probability of each color is 1/4 for each universe, and independent for different universes. This time, let the probability that each universe contains an agent be 1/10. Whether each universe is inhabited is independent of its color and of what the other universes are like.

Again, you know this set-up, and you find yourself in a black room. What probabilities should you assign? Again, we'll take the three rules one by one.

Compartmentalized Conditionalization. Given a single universe, the probability of there being at least one agent in a black room is $1/10 \cdot 1/4 = 1/40 = 0.025$. Given a multiverse, this probability is $1 - (1 - 1/40)^4 \approx 0.096$. So the multiverse hypothesis is relatively confirmed by a factor of ≈ 3.8, which takes its probability from 1/3 to a bit under 0.66.

Self Indication. Given a single universe, the expected number of agents in black rooms is 1/40. Given a four-universe multiverse, this expectation is $4 \cdot (1/40)$. So the multiverse hypothesis is relatively confirmed by a factor of 4, bringing its probability from 1/3 to 2/3.

Self Sampling. This is the most complicated one. Each world containing agents gets weighted according to its proportion of agents in black rooms, and the uninhabited worlds get zero weight. So we weigh each of the rival hypotheses H by the probability it gives to there being at least one agent, multiplied by the expected proportion of agents in black rooms, conditional on H and there being at least one agent. For either hypothesis, this conditional expectation is 1/4, so it can be cancelled out. Given a single universe, the probability of there being at least one agent is the fine-tuning parameter 1/10. Given a multiverse, this probability is $1 - (1 - 1/10)^4 \approx 0.34$. So the multiverse is confirmed by a factor of ≈ 3.4, taking its probability from 1/3 to ≈ 0.63.

Let's sum up (see table 1). Self Indication says that the multiverse gets confirmed to exactly the same degree whether or not life is fine-tuned. Compartmentalized Conditionalization or Self Sampling both say that how much the multiverse gets confirmed is sensitive to fine-tuning. Self Sampling gives no confirmation at all to the multiverse without fine-tuning, but with fine-tuning the multiverse gets a substantial boost. With Compartmentalized Conditionalization, the multiverse is confirmed even without fine-tuning, but in the presence of fine-tuning it is confirmed even more. The general formulas for different parameters in this toy model are also given in table 1.[29]

One more observation. In section 3 we noted that highly specific purely qualitative local evidence strongly confirms a multiverse, and we raised the possibility that self-locating evidence might screen off this confirmation. Now we can see that two of these three rules bear this out. Notice that the formulas for Self Indication and Self Sampling in table 1 do not depend on q: this means that in our simple setting, the specificity of one's mundane evidence makes no difference to how strongly a multiverse is confirmed. (In the toy model, this mundane evidence is "I am in a black room," and q represents the probability that any given agent is in a black room.) So Compartmentalized Conditionalization is the only one of the three rules for which the specificity of mundane evidence gives the multiverse a boost.

6 White on Self-Location

The fact that some universe contains fine-tuned life is powerful evidence in favor of a large multiverse. As we discussed in section 3, additional *local*

[29] How can we derive verdicts from Titelbaum's CLF for these fine-tuning models? The most obvious approach is to use Titelbaum's "technicolor" trick (see 2012, sec. 9.3 and 11.1.2). Consider a variant model where each universe is labeled with a number that can be clearly seen by any agent that inhabits it. In a multiverse, the universes are labeled 1 to n. In a single universe, the universe is assigned a label between 1 and n at random (from a uniform distribution). It seems plausible that these labels do not make a difference to the probability one should assign to being in a multiverse (though we won't argue for this here—and we note that unrestricted Compartmentalized Conditionalization does not agree with this judgment). In this modified case, there is sure to be at most one live center in each world: each universe contains only one center, and centers in different universes that might otherwise have been indistinguishable have been distinguished by being assigned different labels. In this case, CLF says to conditionalize on the evidence "Someone is in a black room in universe k" (for whatever number $1 \leq k \leq n$ is observed). The probability of this holding in a single universe is $1/n \cdot pq$, and the probability of this holding in a multiverse is pq, so the confirmation factor for the multiverse is n, with or without fine-tuning—the same as with Self Indication. (See also discussion of duplication cases in Titelbaum 2012, sec. 11.2.3).

Rule	No fine-tuning	Fine-tuning	Formula
Compartmentalized Conditionalization	≈ 2.7	≈ 3.8	$\frac{1-(1-pq)^n}{pq}$
Self Indication	4	4	n
Self Sampling	1	≈ 3.4	$\frac{1-(1-p)^n}{p}$

Table 1: Relative confirmation factors for the multiverse hypothesis in the two toy models, and for alternative parameter values (given our independence assumption, and assuming that each inhabited universe has at most one agent). Here p is the fine-tuning parameter (the probability of life in a universe), q is the probability that an arbitrary agent is F, where one's self-locating evidence is being F, and n is the number of universes given the multiverse hypothesis. (In the toy model without fine-tuning, $p = 1$, $q = 1/4$, and $n = 4$. In the toy fine-tuning model, $p = 1/10$, $q = 1/4$, and $n = 4$.)

qualitative evidence only piles on additional confirmation. Setting aside the possibility of non-local evidence (and holding fixed the independence assumptions in the background), the only hope for a single universe is *non-qualitative* evidence.

As we discussed in section 2, Hacking (1987) and White (2000) argued that the key question is how the evidence that *our* universe has fine-tuned life makes a difference, over and above the evidence that *some* universe has fine-tuned life. But while they pointed us to the right question, it is harder to answer than either of them acknowledged. Hacking and White each claimed that *our* universe is no more likely to have fine-tuned life given the existence of a multiverse than it is given the existence of a single universe, and thus that the fact that our universe has fine-tuned life does not confirm the existence of a multiverse. But it's just not clear how to think about this.

For a start, while we have helped ourselves to prior probabilities for *qualitative* hypotheses about the number of universes and their intrinsic profiles, it is much more difficult to assess prior probabilities for propositions like *our universe has fine-tuned life*. But this is what we would need to do in order to justify the claim that this is independent of how many universes there are. We can hope to guide qualitative priors to a large extent by physical chances or other measures that arise naturally within cosmology. But it is harder to see how empirical physics might give us much direct guidance about *self-locating* prior probabilities.[30] In fact, the three rules we have discussed give us purchase on claims about *confirmation* by self-locating evidence that does not have to go through claims about prior probabilities of self-locating evidence.

That is not to say we think the project of identifying self-locating priors is hopeless: we take it up in appendix B (see also Arntzenius and Dorr 2017). The way of thinking about things outlined there basically vindicates what White calls the "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, p. 244).[31] Furthermore, each of the self-locating rules we discussed in section 4 can be reinterpreted as recipes for generating self-locating priors from qualitative priors. But as it turns out, none of these particular recipes generate priors that vindicate White's *independence* claim—that whether our universe contains life is independent of how many universes there are.

In a postscript, White (2003) considers more seriously how self-location might make a difference to this issue. Here is what he says:

> It is not enough for confirmation that if my colleagues are rolling dice, it is more likely that *someone* will see a double-six. If my observation is to provide *me* with evidence of these other rolls, they will have to make it more likely that *I* would observe this. . . . What we need is a probabilistic link between my experiences and the hypothesis in question. (White 2003, p. 244, original emphasis)

He then develops this idea in the voice of an interlocutor:

[30]But that might be too pessimistic: many-worlds quantum mechanics might do just that. (For discussion of this interpretive project see, for instance, Greaves 2004). A solution to the "measure problem" in inflationary cosmology might do the same.

[31]That principle is not quite right in cases of introspective failure (though this is not especially important in our present context). There may be cases where one observes oneself to be F, but is not aware of this observation. Let H be "It's raining but I don't observe it." Suppose I observe that it is raining, but don't have this fact about my observation as evidence. My observation that it is raining gives me the evidence that it is raining, which is evidence for H. But the probability that I would *observe* that it is raining, given H, is zero.

The correct principle is that the evidence *I am F* supports hypothesis H only if it is more likely given H that I would *be F* (whether or not I observe this). This principle is a basic fact about conditional probability, given that for E to be evidence for H means that $\Pr(H \mid E) > \Pr(H)$—including the case of self-locating evidence. We can apply this principle to the special case where F is making a certain observation—in the ordinary case where one has evidence *about* one's own observations.

> "... There are very many beings who could have been created other than me. And I'm no more likely to be born in this universe than in any other. 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." (White 2003, p. 244, original emphasis)

This suggestion basically amounts to the Self Indication idea. Before we discuss White's reply, let's pause here. First, we don't object to framing things in terms of a "probabilistic link" between one's self-locating evidence and the hypothesis—though we should beware of taking this too literally, since probabilistic *dependence* in the priors need not involve any kind of causal mechanism. But also, while White doesn't explicitly say this, his framing strongly implies that this Self Indication idea—that one is more likely to observe anything at all if there are more living creatures—is the most promising or the only way of establishing such a probabilistic link. We should be clear that this isn't true. As we have just seen, Compartmentalized Conditionalization and Self Sampling also each provide the requisite "link" between the multiverse and your own observations. (When translated into self-locating priors, in simple models they each make it more likely that you would observe anything if there are many universes, when each universe has a low probability of life.) But neither of these rules builds in the idea that there being more people makes it more likely that you would observe anything. They do each build in the idea that there being at least *one* observer makes it more likely that you would observe anything. But that idea is surely right: after all, there being *zero* observers surely makes it *less* likely that you would observe anything.

Here is how White replies to the Self Indication suggestion:

> The metaphysical picture behind this story is dubious. But, quite apart from that, we can see that something must be wrong with this line of reasoning. The standard argument takes the fact that a universe must be extremely fine-tuned to support life, that a random Big Bang has a very slim chance of producing life, as crucial to the case for multiple universes. If the current objector's argument is cogent, then it should go through regardless of the need for fine-tuning for life. That is, even if a universe with just any set of fundamental constants is bound to produce life, we could still argue along these lines that the more universes there are the more opportunities I had for existing and observing,

> and hence that my observations provide evidence for multiple universes.
>
> Indeed, if the objector's argument is sound, then the discovery that a universe must meet very tight constraints in order to support life should *diminish* the strength of the case for multiple universes. For if every universe is bound to produce life, then by increasing the number of universes we rapidly increase the number of conscious beings, whereas if each universe has a slim chance of producing life, then increasing the number of universes increases the number of conscious beings less rapidly, and hence (by the objector's argument) increases the likelihood of my existence less. I would be surprised if anyone wants to endorse an argument with these consequences, but, at any rate, it is not the standard one that takes the *fine-tuning* data to be crucial in the case for multiple universes. (White 2003, pp. 244–245, original emphasis).

When White says that more stringent fine-tuning would *diminish* the strength of the Self Indication argument for the multiverse, that's a mistake. In our toy model we saw that the confirmation factor was n (the number of universes in the multiverse) regardless of the fine-tuning parameter. White's reasoning only gives half the story: it is correct that the smaller the fine-tuning parameter, the smaller the expected number of conscious beings given a multiverse. But also, in the same way, the smaller the fine-tuning parameter, the smaller the expected number of conscious beings given a *single* universe. (The probability decreases that there will be any conscious beings at all.) These factors cancel out, leaving the confirmation the same for any fine-tuning parameter.

Still, White's more central point in this passage is correct: fine-tuning does not *increase* the strength of confirmation that the Self Indication rule gives to the multiverse. So he's right: the support that Self Indication gives to the multiverse is not really a *fine-tuning* argument. It's a different beast. In that way it comes apart from both of the other rules we considered. Indeed, this feature of Self Indication—that it lends support to huge worlds regardless of the details of our evidence—is one of the major objections it faces.

We should distinguish two ideas. One is that being in a universe that contains fine-tuned life is strong evidence for the multiverse. The second idea is that the *fine-tuning* part of this evidence plays a crucial role, over and above being in a universe that contains life. All three of the rules we have considered support the first conclusion (as will the broader theoretical considerations we offer in section 7). But the second idea is less robust: it is supported by two of the three rules, but not Self Indication. Without doing more to adjudicate between different approaches to self-locating evidence, this is as much as we can say: the question remains pressing.

But at least we can see that if fine-tuning does *not* provide support for the multiverse, it is not for the reason that Hacking and White defended. For their main arguments attack the *first* idea: both of them contend that *our universe contains fine-tuned life* does not provide evidence for the multiverse at all. For the kind of simple multiverse model we have all taken for granted (with its independence assumptions), that evidence *does* support a multiverse, given any reasonable spelled-out theory of self-locating evidence that we know of.

7 Two Multiverse Confirmation Theorems

There are good reasons for dissatisfaction with each of Compartmentalized Conditionalization, Self Indication, and Self Sampling—but it is very hard to come up with a more satisfactory alternative. This topic is hard. In the absence of clarity about the correct self-locating epistemology, we should take a step back and ask whether some other rule might go differently. Might there be some reasonable rule among those we *haven't* thought of that *doesn't* say that, in the presence of fine-tuning, self-locating evidence supports a multiverse? One might hope so. According to Self Indication, the multiverse is no more strongly confirmed with fine-tuning than without fine-tuning. According to Self Sampling, the multiverse is not confirmed without fine-tuning. So one might hope to come up with a single theory that combines both of these features; such a theory would say that a multiverse is not confirmed even with fine-tuning. Is there any reasonable theory like this?

We present two mathematical results that constrain any theory of self-locating evidence—and which strongly suggest that the answer to this question is *no*. We relegate the technical details to appendix A; here we will state the results informally and briefly sketch the main ideas of their proofs. (Appendix B extends these results to the framework of self-locating priors.)

We are now considering abstractly any kind of theory that gives answers to the following kind of question: given certain *prior* probabilities over qualitative hypotheses, and given that you are in a certain qualitative *evidential situation*, what *posterior* probabilities should you assign to qualitative hypotheses? The theories we considered in sections 4 and 5 are examples of such theories, but those theories are quite constrained. For instance, each of those theories assigns the same probability to any two centers in the same qualitative world that are both compatible with one's self-locating evidence.[32] This principle looks suspect in general. For example, in Everettian quantum mechanics, there is a certain physical quantity of "amplitude" or "branch-weight" that ought to make a difference to the epistemic probability of being in one branch of the quantum wave function or another (see for example Greaves 2004; Sebens and Carroll 2016; see also discussion in Titelbaum 2012, pp. 275–276). Two different centers in the same quantum multiverse can each be consistent with one's evidence, but should not be assigned the same probability because they are on branches with different quantum amplitudes. A different kind of challenge comes from conceptions of evidence that do not just separate "live" and "dead" centers, but treat this distinction as a matter of degree (for example, views in the spirit of Jeffrey 1983). There are also many approaches to puzzles about time or memory loss (such as the Sleeping Beauty puzzle), some of which work quite differently from the rules we considered in section 4 (see the references in section 4). Still, all such views that we know of (insofar as they give precise answers to the question in the form we have posed it) are compatible with the abstract framework of this section. Whatever those views might say about other puzzles, the theorems we will present here constrain what they say about fine-tuning and the multiverse.

We suppose there is a *prior* probability distribution over qualitative hypotheses. We consider qualitative priors which have a certain simple form, which builds in the same kind of independence assumptions that Hacking and White supposed, and which we have deployed throughout. Here is the picture. There is a certain probability distribution over possible intrinsic profiles of a universe. However many universes there may be, each universe's intrinsic profile is independent and identically distributed: that is, for each universe u, no matter how many other universes there are, and no matter what their intrinsic profiles are like, the probability of u having a certain intrinsic profile is given by this same distribution. We call priors like this *separable*. The official definition of separability is stated in terms of prior probabilities for purely *qualitative* hypotheses, rather than (as the informal gloss suggested) giving an official place to prior *de re* probabilities

[32]Compare the "highly restricted principle of indifference" of (Elga 2000, p. 144; see also Elga 2004): "Since being in T_1 is subjectively just like being in T_2, and since exactly the same [qualitative] propositions are true whether you are in T_1 or T_2, . . . you ought to have equal credence in each."

about particular universes. This involves some technicalities, discussed in appendix A. The *fine-tuning parameter p* is the probability of there being any centers at all, given a single universe.

This picture is clearly an idealization. For one thing, the multiple universes that appear in contemporary physics are not entirely separate and independent of one another. For another, it could be entirely reasonable to expect different intrinsic properties for an isolated universe than one expects from a universe in a multiverse. For example, one might allocate some prior probability to a theistic hypothesis, according to which God does not care much about how many universes there are, but is adamant that at least one universe contains life. In that case, a universe that is part of a large multiverse might be much less likely to contain life than a single isolated universe. We do not mean to suggest that separable priors are realistic. But seeing how things go in this especially simple idealized case is illuminating.

We also consider an idealizing constraint on *evidence*. So far we have focused on views of evidence according to which it consists in propositions of the form "I am now F," which we have represented by way of properties of agents. But the limiting theorems apply to a much wider range of views about self-locating evidence. For the purposes of these theorems, we can be almost entirely neutral about which features of agents are relevant to the posterior probabilities they should have. We will abstractly consider a relation of *evidential equivalence* between centers: intuitively, this means that the agents of the two centers are exactly alike in whatever respects are relevant to the probabilities they should assign to qualitative hypotheses. This might require that they share phenomenology, or knowledge, or other properties concerning their psychologies, histories, environments, or constitutions.

The main substantive assumption we will make about this relation is that it is *local*, in the following sense. Say two *possible universes* are *evidentially equivalent* iff there is a one-to-one correspondence between the centers in the two universes that takes each center in one universe to an evidentially equivalent center in the other.

Locality. Intrinsic duplicate universes are evidentially equivalent.

For an *internalist* who holds that intrinsic duplicate agents are evidentially equivalent, Locality follows automatically.[33] But Locality is much weaker

[33] This also assumes the modest supervenience principle that any duplication map from one universe to another maps each agent in the first universe to an agent in the other. Intuitively, whether something counts as an agent doesn't depend on anything beyond its universe.

than internalism. The basic idea is that even if evidence "ain't in the head," it is at least "in the universe."

Again, Locality is an idealization. One might well assign some prior probability to non-local evidence. If phenomenal states are evidence, one might assign some prior probability to the hypothesis that the content of one's phenomenal states is partially determined by what goes on in other universes. If knowledge is evidence, one might assign some prior probability to coming to know whether or not there is a multiverse. But it is a reasonable simplification to set such possibilities aside, and it would be surprising if the true theory of multiverse epistemology crucially relied on non-local evidence.

Finally we consider two constraints on *posterior* probabilities. One idea is that confirmation only depends on evidentially relevant features of worlds (whatever those features might be). We say two *worlds* are *evidentially equivalent* iff there is a one-to-one correspondence between the inhabited universes in the two worlds that takes each inhabited universe in one to an evidentially equivalent universe in the other. Given a prior, we call a posterior *evidential* iff no world is confirmed relative to any evidentially equivalent world. The idea is that if something makes a difference to relative probabilities, it should show up somewhere in a world's distribution of "evidential situations", however we are thinking about that notion. (As usual, when we say "H_1 is confirmed relative to H_2," we mean that the ratio of the posterior probability of H_1 to the posterior probability of H_2 is greater than their ratio of prior probabilities.)

Finally, we only consider agents who can tell that there is at least one agent.

Agents. The posterior probability that there are no agents is zero.

There are two main results. The first concerns "small enough" multiverses.

Theorem 1. For any separable qualitative prior and any evidential posterior, Locality and Agents imply that, for any possible multiverse size $1 < n < 1/p$, where p is the fine-tuning parameter, the proposition *there are n universes* is confirmed relative to *there is just one universe*.

Proof Sketch. We'll illustrate with the case $n = 2$. In this case, the hypothesis is that the fine-tuning parameter $p < 1/2$. The main idea of the proof is to consider how things go if we know exactly one universe is inhabited. Let

U_n be the proposition *there are exactly n universes*. Let I_1 be the proposition *exactly one universe is inhabited*.

First, for a separable prior $\Pr(-)$, a little calculation shows that if $p < 1/2$, then it is more probable that exactly one universe is inhabited if there are *two* universes than if there is just *one*. (For two universes this probability is $2p(1-p)$, while for one universe it is p.)

Second, we have a lemma that says that, given the number of *inhabited* universes, our self-locating evidence tells us nothing about how many *uninhabited* universes there are (lemma 5 in appendix A). So, conditional on exactly one universe being inhabited, neither U_1 nor U_2 is confirmed relative to the other.

Third, on the posterior we know that if there is just one universe, then it is inhabited. So $U_1 I_1$ has the same posterior probability as U_1, while $U_2 I_1$ is no more probable than U_2.

Put these three steps together, and we're done:

$$\frac{\Pr(U_2)}{\Pr(U_1)} < \frac{\Pr(U_2 I_1)}{\Pr(U_1 I_1)} = \frac{\Pr_*(U_2 I_1)}{\Pr_*(U_1 I_1)} \leq \frac{\Pr_*(U_2)}{\Pr_*(U_1)}$$

That is, U_2 is confirmed relative to U_1. □

Theorem 1 tells us that for any n, there is some fine-tuning parameter p for which a size-n multiverse is confirmed. So a multiverse of any finite size is confirmed by fine-tuned life—as long as it is fine-tuned *enough*.

Our second result generalizes to arbitrarily large multiverses; but it relies on a principle whose status does not seem nearly as clear as the others. The idea is that *without* any fine-tuning, our evidence does not *disconfirm* a multiverse. We say a posterior is *ordinary* iff, for each $n > 0$, conditional on every universe containing agents, the existence of just *one* universe is not confirmed relative to the existence of n universes.

Posteriors that are not ordinary seem a bit odd (in a context of separable priors and local evidence). If we suppose that every universe has life, but each universe's internal configuration is independent of any other's, and our evidence only directly tells us about our universe's internal configuration, then it's not clear why we would think that our evidence favors a single universe. That's not to say it couldn't turn out this way. Indeed, this *can* happen with Self Sampling. If our evidence indicates that our universe has a relatively *small* population, compared to the expected population size of an arbitrary inhabited universe, then the expected proportion of centers compatible with our evidence is higher for a single universe than it is for a multiverse. So this kind of evidence would favor a single universe over a multiverse, by Self Sampling. (Perhaps we even have this kind of evidence: looking around, one might take our own universe to be surprisingly sparsely inhabited, compared to what one might expect from an inhabited universe.)[34]

We put forward "ordinariness" in a different spirit from the other constraints we have discussed. It is not a very plausible constraint on what posteriors *have* to be like in general. It is also not really a simplifying idealization like separability or locality. But theorem 2 tells us something interesting: given the other assumptions, the *only* way that a large multiverse can fail to be confirmed in the presence of fine-tuning is if our posteriors are "extraordinary," which is to say (roughly) that in the *absence* of fine-tuning a multiverse is *disconfirmed*. (Furthermore, the proof of the theorem shows us that for large n and small p, the only way to avoid confirming a multiverse is if, in the absence of fine-tuning, our evidence *strongly* disconfirms a multiverse.)

This again tells against Hacking and White's diagnosis of the fine-tuning argument. If they were right, then our evidence should not count *against* a multiverse, either—even in the absence of fine-tuning. (If it is clear that this roll of the dice coming up double sixes does not support there being many rolls, it is even clearer that, conditional on every roll coming up double sixes, this roll of the dice coming up double sixes does not support there being just *one* roll of the dice!)

Theorem 2. For any separable prior and any *ordinary* evidential posterior, Locality and Agents imply that for any possible multiverse size $n > 1$, *there are n universes* is confirmed relative to *there is just one universe*.

The proof uses similar ideas to theorem 1. For each $1 \leq k \leq n$, we can show that *there are n universes exactly k of which are inhabited* is not disconfirmed relative to *there is one inhabited universe* (for separable priors, local evidence, and ordinary evidential posteriors). It follows that *there are n universes and at least one universe is inhabited* is not disconfirmed relative to *there is one inhabited universe*. But the *prior* probability that at least one universe is inhabited is higher if there are many universes than if there is just one. For the rest of the argument see appendix A.

[34] The argument of Olum (2004) is closely related.

8 Conclusion

The existence of fine-tuned life, taken on its own, is straightforwardly evidence for a multiverse. But this mere qualitative evidence should *not* be taken on its own: our relevant evidence is not just that there is some fine-tuned life or other, but that *we ourselves* are fine-tuned life. This can make a difference. We have also evaluated a seductive argument that this more specific evidence does *not* confirm the multiverse. Some standard rules for self-locating probabilities do not vindicate this idea. What's more, our general results show that it will be difficult to come up with any reasonable theory that does vindicate it. If being in a universe with fine-tuned life is not evidence for the existence of a multiverse, then some principle in the multiverse confirmation theorems must fail: in the case of "small" multiverses, this means that priors are not separable, evidence is not local, or posteriors are not evidential. This is not to end the debate, but it provides a framework in which to continue it. A duel of conflicting intuitive analogies won't cut it: clear reasoning about the strength of our evidence for a multiverse requires systematic theorizing about self-locating evidence.

A Theorems and Proofs

Here we present precise statements and proofs of the two theorems in section 7.

Definition 1. A **multiverse structure** consists of the following components.

1. A non-empty set W of **possible worlds**.
2. A non-empty set I of **centers**.
3. A non-empty set U of **(possible) universes**. Each universe is **in** exactly one world, and each center is **in** exactly one universe.
4. An equivalence relation between universes of **intrinsic duplication**.
5. An equivalence relation between centers of **evidential equivalence.**

In what follows we restrict attention to the case where W and I are both countable. (This relaxes the assumption in section 4 of only finitely many worlds or centers.) We are thinking of worlds as specified qualitatively. A **(qualitative) proposition** is a set of worlds.

(In supposing that every center is in a universe, we set aside the a priori possibility of being an immaterial agent like an angel or a god. This is another idealization.)

Definition 2. For any multiverse structure,

(a) A universe u is **inhabited** iff there is at least one center in u.

(b) For any number n, U_n is the proposition that there are exactly n universes (that is, the set of worlds w such that there are exactly n universes in w).

(c) For any number n, I_n is the proposition that there are exactly n inhabited universes.

We can use the evidential equivalence relation on *centers* to define closely related relations on universes and worlds.

Definition 3.

(a) *Universes* u and u' are **evidentially equivalent** iff there is a one-to-one correspondence f from the centers in u to the centers in u' such that i is evidentially equivalent to $f(i)$ for each center i in u.

(b) *Worlds* w and w' are **evidentially equivalent** iff there is a one-to-one correspondence f from the inhabited universes in w to the inhabited universes in w' such that u is evidentially equivalent to $f(u)$ for each universe u in w.

(c) A multiverse structure is **local** iff any pair of intrinsic duplicate universes are evidentially equivalent.

Let a **qualitative prior** $\Pr(-)$ on a multiverse structure be a (countably additive) probability function on the set of all qualitative propositions. We will only consider priors with a certain simple structure. The intuitive idea is that there is a distribution over possible intrinsic profiles of a universe, and that each universe is assigned an intrinsic profile at random from this distribution: so the profiles of all of the universes in a world are independent and identically distributed. This is a bit difficult to state officially in purely qualitative terms: the most natural way of doing it would presuppose "trans-world identity" for possible universes, which our model does not build in. But there is a fairly straightforward way of describing the qualitative probability distribution that corresponds to this picture.

A first pass at the idea is that the probability of any particular qualitative profile of a multiverse, given its number of universes, is the product of the probabilities of the various individual universe-profiles that make it up. This isn't quite right, though, because sometimes there is more than one way to get a qualitative profile. If you roll two fair dice, the probability of rolling a five and a six is twice the probability of two sixes. While there is only one way to roll two sixes, there are two ways to get a five and six: the pairs $(5, 6)$ and $(6, 5)$. Independent universe profiles should have the same structure. In general, the number of different ways to get a sequence of n things, which includes d distinct elements repeated $k_1, \dots, k_d$ times, respectively, is

$$\frac{n!}{k_1! \cdots k_d!}$$

This is the number of permutations of n indices, with permutations that take elements to indistinguishable elements quotiented out.

For a sequence of possible universes $(u_1, \dots u_n)$, let the **local profile** $[u_1 \cdots u_n]$ be the set of worlds w such that there are distinct universes $v_1, \dots v_n$ in w which are duplicates of $u_1, \dots, u_n$, respectively. Intuitively this proposition says that the multiverse includes a copy of a $(u_1, \dots, u_n)$ multiverse.

Definition 4. Call a qualitative prior $\Pr(-)$ on a multiverse frame **separable** iff it satisfies the following three conditions.

(a) We say n is a **possible multiverse size** iff $\Pr(U_n) > 0$. For any numbers $1 \leq k \leq n$, if n is a possible multiverse size, then k is a possible multiverse size as well.[35]

(b) Let n be any possible multiverse size, and let $u_1, \dots, u_n$ be a sequence of possible universes. Suppose that this sequence represents d different equivalence classes under duplication, which are repeated $k_1, \dots, k_d$ times in the sequence respectively. Then:

$$\Pr([u_1, \dots, u_n] \mid U_n) = \frac{n!}{k_1! \cdots k_d!} \Pr([u_1] \mid U_1) \cdots \Pr([u_n] \mid U_1)$$

(c) The **fine-tuning parameter** $p = \Pr(I_1 \mid U_1)$ satisfies $0 < p < 1$.

[35]The main purpose of this "regularity" condition is to avoid division by zero, especially in lemma 1 (a) and the proof of theorem 2. For the purposes of theorem 1, this assumption could be eliminated at the cost of some technical inconvenience.

Here are some consequences of this structure. We omit the proofs, which involve some simple combinatorics.

Lemma 1. *Let* $\Pr(-)$ *be a separable qualitative prior; let* n *be a possible multiverse size; let* p *be the fine-tuning parameter.*

(a) Suppose that none of the universes $u_1, \ldots, u_k$ *are duplicates of any of the universes* $u_{k+1}, \ldots, u_n$*. Then*

$$\Pr([u_1 \cdots u_n] \mid U_n) = \binom{n}{k} \Pr([u_1 \cdots u_k] \mid U_k) \Pr([u_{k+1} \cdots u_n] \mid U_{n-k})$$

where $\binom{n}{k}$ *is the binomial coefficient* $\frac{n!}{k!\,(n-k)!}$*.*

(b) For $0 \le k \le n$*,*

$$\Pr(I_k \mid U_n) = \binom{n}{k} p^k (1-p)^{n-k}$$

In particular, given that $0 < p < 1$*,* $\Pr(I_k \mid U_n) > 0$

(c) If $1 < n$*, it is more likely that there is at least one inhabited universe if there are* n *universes than if there is just one:*

$$\Pr(\neg I_0 \mid U_n) > \Pr(\neg I_0 \mid U_1)$$

(d) If $1 < n < 1/p$*, it is more likely that there is exactly one inhabited universe if there are* n *universes than if there is just one:*

$$\Pr(I_1 \mid U_n) > \Pr(I_1 \mid U_1)$$

Lemma 2. *Let* $\Pr(-)$ *be a separable qualitative prior on a local multiverse structure. Let* $0 \le k \le n$*, where* n *is a possible multiverse size. Let* $(u_1, \ldots, u_k)$ *be a sequence of inhabited universes, and let* Q *be the local profile* $[u_1 \cdots u_k]$*. Then*

$$\Pr(Q \mid U_n\, I_k) = \Pr(Q \mid U_k I_k)$$

In particular, this does not depend on n*.*

Proof. We have

$$\Pr(U_n\, I_k\, Q) = \sum_R \Pr(U_n\, Q\, R)$$

where R ranges over the propositions $[v_1 \cdots v_{n-k}]$ for each sequence of uninhabited universes $v_1, \ldots, v_{n-k}$. This is equal to

$$
\begin{aligned}
&\Pr(U_n) \sum_R \Pr(Q\, R \mid U_n) \\
&= \Pr(U_n) \binom{n}{k} \Pr(Q \mid U_k) \sum_R \Pr(R \mid U_{n-k}) && \text{lemma 1 (a)} \\
&= \Pr(U_n) \binom{n}{k} \Pr(Q \mid U_k) \Pr(I_0 \mid U_{n-k}) \\
&= \Pr(U_n) \binom{n}{k} \Pr(Q \mid U_k) \, (1-p)^{n-k} && \text{lemma 1 (b)}
\end{aligned}
$$

where p is the fine-tuning parameter. Also, by lemma 1 (b),

$$\Pr(U_n\, I_k) = \Pr(U_n) \binom{n}{k} p^k (1-p)^{n-k}$$

The conditional probability $\Pr(Q \mid U_n\, I_k)$ is the quotient of these two quantities, namely:

$$\frac{\Pr(Q \mid U_k)}{p^k}$$

This is equal to

$$\frac{\Pr(U_k Q)}{\Pr(U_k) \cdot p^k} = \frac{\Pr(U_k Q)}{\Pr(U_k I_k)} = \frac{\Pr(U_k I_k Q)}{\Pr(U_k I_k)} = \Pr(Q \mid U_k I_k)$$

The second equality holds because Q is the local profile for k inhabited universes, so $U_k I_k Q$ and $U_k Q$ are equivalent. □

Next we will consider a *posterior* probability distribution $\Pr_*(-)$, also defined on the same space of worlds.

Definition 5. For a prior $\Pr(-)$ and a posterior $\Pr_*(-)$, and for propositions A and B, we say that A **is confirmed relative to** B iff

$$\Pr_*(A) \Pr(B) > \Pr_*(B) \Pr(A)$$

We write $A \succ_* B$ for this relation. Similarly, we say $A \succsim_* B$ iff B is not confirmed relative to A, and $A \sim_* B$ iff neither A nor B is confirmed relative to the other.

We say A **is confirmed relative to** B **conditional on** C iff $A\, C$ is confirmed relative to $B\, C$.

It is a bit more transparent what relative confirmation means if we write it in this form:

$$\frac{\Pr_*(A)}{\Pr_*(B)} > \frac{\Pr(A)}{\Pr(B)}$$

This says that the ratio of the posterior probabilities of A and B is higher than the ratio of their priors. But the ratio formulation has the disadvantage of requiring us to be ever-vigilant about division by zero.

The following basic facts about confirmation follow from the probability calculus. We omit the proofs.

Lemma 3. *Consider any qualitative prior* $\Pr(-)$ *and posterior* $\Pr_*(-)$*, and any propositions* A*,* B*, and* C*.*

(a) *If* $\Pr(C) = 0$*, then* A *and* B *are not relatively confirmed conditional on* C*: that is,* $A\,C \sim_* B\,C$*.*

(b) *If* $\Pr(B) \neq 0$*,* $A \gtrsim_* B$*, and* $B \gtrsim_* C$ *then* $A \gtrsim_* C$*.*

(c) *Let* $\mathcal{E}$ *be a set of propositions. We say* $\mathcal{E}$ ***partitions*** A *iff every world in* A *is in exactly one* $E \in \mathcal{E}$*. If* $\mathcal{E}$ *is a countable set that partitions* A*, and* $A\,E \gtrsim_* B$ *for each* $E \in \mathcal{E}$*, then* $A \gtrsim_* B$*. Likewise, if* $B \gtrsim_* A\,E$ *for each* $E \in \mathcal{E}$*, then* $B \gtrsim_* A$*.*

(d) *Let* $\mathcal{E}$ *be a countable set of propositions that partitions both* A *and* B*. If*

$$\Pr_*(A\,E)\ \Pr(B) \geq \Pr_*(B\,E)\ \Pr(A) \quad \text{for each } E \in \mathcal{E}$$

then $A \gtrsim_* B$*.*

We consider two constraints on posteriors. First, one's evidence rules out there being no agents at all.

Agents. $\Pr_*(I_0) = 0$

The second idea is that if two worlds are alike in every evidentially relevant respect, then neither is confirmed over the other.

Definition 6. Let $\Pr(-)$ be a qualitative prior. We call a posterior $\Pr_*(-)$ **evidential** iff no pair of evidentially equivalent worlds is relatively confirmed: that is, for any evidentially equivalent worlds w_1 and w_2, $w_1 \sim_* w_2$.

We can immediately rewrite this in a more general form.

Lemma 4. *For any qualitative prior and any evidential posterior, if E is a set of pairwise evidentially equivalent worlds, for any propositions A and B, AE is not confirmed relative to BE; or in other words, A is not confirmed relative to B conditional on E.*

Proof. If $\Pr(E) = 0$ then we are done by lemma 3 (a). Otherwise, there is some world $w \in E$ such that $\Pr(w) > 0$. (Here we use our assumption that the set of worlds W is countable.) The proposition $A\,E$ is partitioned by the set of worlds $v \in A\,E$. Each world $v \in A\,E$ is evidentially equivalent to w, so we know that $v \sim_* w$. By lemma 3 (c), $A\,E \sim_* w$. By parallel reasoning, $w \sim_* B\,E$. So $A\,E \sim_* B\,E$ by lemma 3 (b). ◻

Lemma 5. *Let* $\Pr(-)$ *be a separable qualitative prior on a local multiverse structure, and let* $\Pr_*(-)$ *be an evidential posterior. For any possible multiverse sizes m and n, U_n is not confirmed relative to U_m conditional on I_k.*

Proof. Let $\mathcal{Q}$ be the set of distinct local profiles $Q = [u_1 \cdots u_k]$ where $(u_1, \dots, u_k)$ is a sequence of inhabited universes and $\Pr(Q) > 0$. These propositions are mutually exclusive.

Consider any $Q \in \mathcal{Q}$. Any two worlds in $I_k\,Q$ are evidentially equivalent: each of the k inhabited universes in one world can be mapped to a duplicate universe in the other, and locality says that duplicate universes are evidentially equivalent. Thus lemma 4 tells us that neither of U_n or U_m is relatively confirmed conditional on $I_k\,Q$. So we have

$$\Pr_*(U_n\,I_k\,Q) \cdot \Pr(U_m\,I_k\,Q) = \Pr_*(U_m\,I_k\,Q) \cdot \Pr(U_n\,I_k\,Q)$$

Lemma 2 tells us:

$$\Pr(Q \mid U_n\,I_k) = \Pr(Q \mid U_m\,I_k)$$

And so:

$$\Pr(U_n\,I_k\,Q) \cdot \Pr(U_m\,I_k) = \Pr(U_m\,I_k\,Q) \cdot \Pr(U_n\,I_k)$$

Multiplying these two equations together and canceling the (non-zero) factor $\Pr(U_n\,I_k\,Q) \cdot \Pr(U_m\,I_k\,Q)$ we find:

$$\Pr_*(U_n\,I_k\,Q) \cdot \Pr(U_m\,I_k) = \Pr_*(U_m\,I_k\,Q) \cdot \Pr(U_n\,I_k)$$

The conclusion follows by lemma 3 (d). ◻

Theorem 1. *Let* $\Pr(-)$ *be a separable qualitative prior on a local multiverse structure, and let* $\Pr_*(-)$ *be an evidential posterior that satsfies Agents. For any possible multiverse size n, if $1 < n < 1/p$, where p is the fine-tuning parameter, then U_n is confirmed relative to U_1.*

Proof. If $\Pr_*(U_1) = 0$ then we're done; so suppose $\Pr_*(U_1) > 0$. By Agents (and the fact that U_1 entails $I_0 \vee I_1$) we have $\Pr_*(U_1\, I_1) = \Pr_*(U_1) > 0$. So

$$\begin{aligned}
\frac{\Pr_*(U_n)}{\Pr_*(U_1)} &= \frac{\Pr_*(U_n)}{\Pr_*(U_1\, I_1)} \\
&\geq \frac{\Pr_*(U_n\, I_1)}{\Pr_*(U_1\, I_1)} \\
&= \frac{\Pr(U_n\, I_1)}{\Pr(U_1\, I_1)} && \text{lemma 5} \\
&> \frac{\Pr(U_n)}{\Pr(U_1)} && \text{lemma 1 (d)} \qquad \square
\end{aligned}$$

The second result uses the idea is that one's evidence does not count *against* a multiverse, conditional on there being *no* fine-tuning for life.

Definition 7. Given a qualitative prior $\Pr(-)$ on a multiverse structure, say a posterior $\Pr_*(-)$ is **ordinary** iff for any possible multiverse size n, U_1 is not confirmed relative to U_n conditional on every universe being inhabited. That is, a posterior $\Pr_*(-)$ is ordinary iff for each n such that $\Pr(U_n) > 0$,

$$U_n\, I_n \gtrsim_* U_1\, I_1$$

Theorem 2. *Let* $\Pr(-)$ *be a separable qualitative prior on a local multiverse structure. Let* $\Pr_*(-)$ *be an ordinary evidential posterior that satisfies Agents. For any possible multiverse size* $n > 1$, U_n *is confirmed relative to* U_1.

Proof. Suppose $1 \leq k \leq n$. The conditions on a separable prior tell us that $\Pr(U_k\, I_k) > 0$ So by lemma 5, ordinariness, and lemma 3 (b),

$$U_n\, I_k \sim_* U_k\, I_k \gtrsim_* U_1\, I_1$$

The propositions $I_1, \dots, I_n$ partition $U_n\, \neg I_0$, so by lemma 3 (c),

$$U_n\, \neg I_0 \gtrsim_* U_1\, I_1$$

Finally,

$$\begin{aligned}
\frac{\Pr_*(U_n)}{\Pr_*(U_1)} &= \frac{\Pr_*(U_n\, \neg I_0)}{\Pr_*(U_1\, I_1)} && \text{by Agents} \\
&\geq \frac{\Pr(U_n\, \neg I_0)}{\Pr(U_1\, I_1)} && \text{as shown above} \\
&= \frac{\Pr(\neg I_0 \mid U_n)\, \Pr(U_n)}{\Pr(\neg I_0 \mid U_1)\, \Pr(U_1)} \\
&> \frac{\Pr(U_n)}{\Pr(U_1)} && \text{lemma 1 (c)} \qquad \square
\end{aligned}$$

The proof also gives us a lower bound on the *strength* of confirmation. For separable priors, the probability $\Pr(\neg I_0 \mid U_n)$ of at least one universe being inhabited if there are n universes total is $1 - (1-p)^n$, where p is the fine-tuning parameter.

Corollary 1. *Let* $\Pr(-)$ *be a separable prior, and let* $\Pr_*(-)$ *be an ordinary evidential posterior that satisfies Agents. For any possible multiverse size* n*, either* $\Pr_*(U_1) = 0$ *or else*

$$\frac{\Pr_*(U_n)}{\Pr_*(U_1)} \geq \frac{\Pr(U_n)}{\Pr(U_1)} \left(\frac{1-(1-p)^n}{p} \right)$$

where p *is the fine-tuning parameter. For large* n*, this confirmation factor approaches* $1/p$*.*

In other words, for separable priors and *ordinary* evidential posteriors, self-locating evidence confirms a multiverse at least as strongly as the qualitative evidence "There is life" does.

B Self-Locating Priors

In appendix A we only considered prior and posterior probabilities for qualitative hypotheses. It's worth thinking about how we can apply a more traditional Bayesian model. If there are also *self-locating prior probabilities*—not just prior probabilities of the qualitative world being a certain way, but prior probabilities of you yourself being a certain way—then updating can proceed in the traditional way: by *conditionalizing* the self-locating prior on one's self-locating total evidence. These priors need not represent any person's primordial credences (no more than the qualitative priors do); rather they would be "ur-priors" that encode facts about evidential support (see Meacham 2016, and references therein; Arntzenius and Dorr 2017). The difficulty is in figuring out what those self-locating prior probabilities should look like.

First a few preliminaries. While in appendix A we could afford to be extremely non-committal about the nature of self-locating evidence, for this section we take the more opinionated stand that one's self-locating evidence is the sort of thing that can be conditionalized on. As we discussed in section 4, we can theorize about a proposition expressed by a sentence "I am now F" by way of the property of being F, which we can represent as a set of centers. To make this precise, we deploy some standard tools from epistemic logic—with *centers* taking center stage, rather than worlds.

Definition 8. A **(centered) frame** consists of the following:

1. A non-empty set W of **worlds**.
2. A non-empty set I of **centers**. Each center is **in** exactly one world.
3. A relation of **evidential accessibility** between centers. We write $E(i)$ for the set of centers that are accessible from i.

What it intuitively means for a center j to be evidentially accessible from a center i is that one's self-locating evidence at i is consistent with being at j.[36] To put that another way, we can consider the conjunction of all properties *being F* which are part of agent a's self-locating evidence at time t in world w. This property is represented by a set of centers, which is the set $E(i)$ where i is the center $\langle a, t, w \rangle$.

It is common to take for granted that a center j is *compatible* with the evidence at i if and only if one has the *same evidence* at j as at i, so i and j rule out precisely the same centers. This amounts to supposing that evidential accessibility is an equivalence relation, which amounts to a strong *access* principle to the effect that one can always tell what one's evidence is (see for example Williamson 2000, ch. 10; Salow 2018) But this assumption is highly contentious, and unnecessary for the results that follow; we do not make it.[37]

Even though we are not assuming that evidential accessibility is an equivalence relation, we can use it to define a closely related equivalence relation:

- Centers i and j are **evidentially equivalent** iff $E(i) = E(j)$ and for any center k, $i \in E(k)$ iff $j \in E(k)$.

This is the weakest equivalence relation that is extended by the accessibility relation. Thus we can let a **multiverse frame** be a multiverse structure in the sense of definition 1, together with an accessibility relation on centers,

[36]To be more explicit, if $i = \langle a, t, w \rangle$ and $j = \langle a', t', w' \rangle$, this means that being the qualitative way a is at t in w implies having only evidence that is compatible with being the qualitative way a' is at t' in w'.

[37]For example, some contend that an agent's evidence is what they know (Williamson 2000, ch. 9). Moreover, it's commonplace to say that an agent with hands knows that they are not a handless brain-in-a-vat with misleading appearances, and that a handless brain-in-a-vat with misleading experiences does not know whether they are an agent with hands or a handless brain-in-a-vat with misleading appearances. Then the access principle would fail. Skeptical hypotheses like these are particularly poignant in the context of cosmology: if the world may well be vast, the chance of there really being many agents in unfortunate skeptical predicaments—like "Boltzmann brains"—is not negligible.

where the structure's evidential equivalence relation is defined in terms of accessibility in this way.

Now let's think about self-locating priors. Consider a simple case. There are three cards, reading *Zero*, *One*, and *Two*. One of them is drawn by a fair chance process. If Zero is drawn, there will be no agents; if One, then there will be exactly one person in a black room; if Two, there will be exactly two people in black rooms. It is straightforward enough that each of these three *qualitative* scenarios should have prior 1/3. But self-locating prior probabilities are much murkier. It isn't clear how to come up with reasonable prior probabilities for propositions like that expressed by "There is exactly one person, and it's *me*." No objective chance was specified for this, and it's a bit mysterious how to even think about it—now we also need to assign some prior probability to the alternative proposition expressed by "There is exactly one person, and it *isn't* me."

The way forward is to take this seriously: we can assign prior probabilities to not being a person or an agent at all. This might have been so—indeed, there might have been no one at all. And the fact that there *are* agents—and that you are one of them—can do significant evidential work, supporting some hypotheses over others. The ur-prior encodes this kind of potential to do evidential work.

There are two different ways of thinking about what it is like not to be an agent: *contingentism* and *necessitism*. The contingentist holds that one might have been nothing whatsoever—and so not an agent. The necessitist holds that, come what may, one would have been *something*—but this something might have been a very boring non-concrete object, without any thoughts or feelings or spatiotemporal location. In either case, one would not have been an *agent*. We will describe possibilities in these neutral terms, rather than using more loaded language about whether one would have *existed*. (To be clear, the relevant modality here is prior *epistemic* possibility, rather than metaphysical possibility. But the structural issue is the same either way.)

The self-locating prior approach requires an innovation (borrowed from familiar models for free logics). In addition to the "agent-centers", in each world there is one additional *null center*, which represents the possibility of *not* being an agent.[38] We suppose that the evidential accessibility relation is defined only on non-null centers: substantively, this is to assume that one

[38]We use just one null center per world because it is simple and sufficient for our application, but other applications might call for assigning each world an "outer domain" that contains many distinct "null centers." Null centers need not be anything exotic: they might just be triples $\langle x, t, w \rangle$ where x is *not* an agent.

is sure to have evidence that entails being an agent, if one is an agent at all. (This is another idealization.)

Just as we modeled self-locating *evidence* as a set of centers, we can extend this to self-locating prior and posterior probabilities as well. A **self-locating prior** $\Pr(-)$ is a probability measure defined for all sets of centers, representing the prior probability of being one way or another.

In appendix A we modeled *qualitative* propositions as sets of worlds. We can straightforwardly associate each set of worlds X with a *boring* set of centers: namely, those centers that are in some world in X (including the null centers). Given a *self-locating* prior $\Pr(-)$, we can thus read off a *qualitative* prior $\Pr_q(-)$ by considering the probabilities of boring sets of centers.

Among the properties we are interested in are properties that do not entail being an agent. The set of centers that represents such a property will include some null centers. This is important, since we want to assign non-zero prior probabilities to *qualitative propositions* like *there are no agents*, and thus also to qualitative properties like *being such that there are no agents* or more simply *not being an agent*. No agent in any world has either of these properties. The positive prior probability that properties like these receive can only come from the positive prior probabilities of null centers.

In this setting, we can get self-locating *posterior* probabilities by just conditionalizing on one's self-locating evidence.

Definition 9. Let $\Pr(-)$ be a self-locating prior on a centered frame. For any (non-null) center i such that $\Pr(E(i)) > 0$, we define the **self-locating posterior**

$$\Pr_i(-) = \Pr(- \mid E(i))$$

(where as before $E(i)$ is the set of centers that are evidentially accessible from i).

We can reconstrue rules like Compartmentalized Conditionalization, Self Indication, and Self Sampling in this framework. These rules took for granted that the posterior is determined by a *qualitative* prior together with indexical evidence. In this framework, such rules can be reinterpreted as rules for mapping a *qualitative* prior $\Pr_q(-)$ to an *indexical* prior $\Pr(-)$ that *enriches* $\Pr_q(-)$. (That is, $\Pr(-)$ assigns the same probability to each boring set of centers as $\Pr_q(-)$ assigns to its associated set of worlds.) The self-locating posterior is then given by conditionalization.

For simplicity we restrict attention here to the case where there are only finitely many worlds containing finitely many centers.[39] In order to specify enriched self-locating priors, it's enough to specify the conditional probability $\Pr(i \mid w)$ for each center i in world w. (Here i and w stand for the obvious corresponding sets of centers.) The probability of an arbitrary qualitative property P can then be calculated as

$$\Pr(P) = \sum_{\substack{w \in W \\ i \in P}} \Pr(i \mid w) \Pr_q(w)$$

Furthermore, it suffices to define $\Pr(i \mid w)$ for each *non-null* center i in a world w. We just choose $\Pr(j \mid w)$ for the null center j so that the conditional probabilities add up to one. Here are the recipes for generating self-locating priors corresponding to each of the three rules we have discussed.

Self Sampling. For any non-null center i in world w,

$$\Pr(i \mid w) = \frac{1}{n(w)}$$

where $n(w)$ is the number of non-null centers in w.

The prior probability of each world containing non-null centers is evenly divided among those centers. Worlds that only have a null center put all their probability there.

Self Indication. For any non-null center i in world w,

$$\Pr(i \mid w) = \frac{1}{N}$$

where N is the maximum number of non-null centers in any world.

The idea is that when there are fewer agents in a world, there is more prior probability left over for the null center.

Compartmentalized Conditionalization. This case is more delicate: there is a formula that works in cases where the structure of possible evidence is especially simple, but not in general. It suffices for the evidential accessibility relation to be an equivalence relation. In that

[39]In fact, even if there are infinitely many worlds, the same formulas will work as long as the number of centers per world is *bounded*. (For Compartmentalized Conditionalization, let T be the maximum number of evidential types in any world, instead of the total number of evidential types.) The Self-Sampling recipe continues to work even if the number of centers per world is *finite but unbounded*.

case, each non-null center i has a unique *evidential type*: there is a unique set of centers $E(j)$ (for any center j) that includes i. Let $t(i)$ be the number of centers in i's world that are of the same evidential type as i.[40] Let T be the total number of distinct evidential types. For any non-null center i in world w,

$$\Pr(i \mid w) = \frac{1}{t(i)\,T}$$

The idea is that the prior probability of a world is first evenly divided among all evidential types, and then each type's probability is evenly divided again among its centers in that world. The probability of a type that has no centers in a world all goes to the null center.

These indexical priors agree with the original rules, in the following sense. Start with any qualitative prior $\Pr_q(-)$. If you generate a self-locating prior using the rule's recipe above, then you conditionalize the prior on the self-locating evidence for a center i (where $E(i)$ has positive prior probability), you get the same result as the original rule gives for i. We omit the proofs.

Since each self-locating prior straightforwardly generates corresponding qualitative priors and posteriors (via boring sets of centers), the constraints on priors and posteriors we stated in appendix A can be applied straightforwardly in this setting without modification. But since now we are requiring posteriors to be generated by conditionalization, we can simplify the constraints a little. Here is a new principle.

Definition 10. Call a self-locating prior $\Pr(-)$ **evidentially natural** iff, for any evidentially equivalent worlds w_1 and w_1, where the equivalence mapping takes center i_1 in w_1 to the center i_2 in w_2, the probability of being at center i_1 given that one is in world w_1 is the same as the probability of being at center i_2 given that one is in world w_2:

$$\Pr(i_1 \mid w_1) = \Pr(i_2 \mid w_2)$$

if $\Pr(w_1) > 0$ and $\Pr(w_2) > 0$.

Lemma 6. *For any evidentially natural self-locating prior* $\Pr(-)$*, for any center i such that* $\Pr(E(i)) > 0$*, the posterior* $\Pr_i(-)$ *is evidential (definition 6).*

[40] The necessary and sufficient condition for Compartmentalized Conditionalization to be representable by a self-locating prior is that $t(i)$ is well-defined: that is, all evidential types that contain a common center i must have the same number of elements in i's world.

Proof. Let w_1 and w_2 be evidentially equivalent worlds. The equivalence map takes each center j in $w_1 \cap E(i)$ to a center $f(j)$ in $w_2 \cap E(i)$ (where w_1 and w_2 stand for the obvious boring sets of centers). The case where $\Pr(w_1) = 0$ or $\Pr(w_2) = 0$ is easy, so assume otherwise. In that case:

$$\begin{aligned}
&\Pr_i(w_1) \cdot \Pr(w_2) && \\
&= \frac{\Pr(w_1 \cap E(i)) \cdot \Pr(w_2)}{\Pr(E(i))} && \text{definition of } \Pr_i(-) \\
&= \sum_{j \in w_1 \cap E(i)} \frac{\Pr(j) \cdot \Pr(w_2)}{\Pr(E(i))} && \\
&= \sum_{j \in w_1 \cap E(i)} \frac{\Pr(j \mid w_1) \cdot \Pr(w_1) \cdot \Pr(w_2)}{\Pr(E(i))} && \\
&= \sum_{k \in w_2 \cap E(i)} \frac{\Pr(k \mid w_2) \cdot \Pr(w_1) \cdot \Pr(w_2)}{\Pr(E(i))} && \text{evidential naturalness} \\
&= \Pr_i(w_2) \cdot \Pr(w_1) && \text{by parallel reasoning}
\end{aligned}$$

□

Theorem 3. *Let* $\Pr(-)$ *be a separable, evidentially natural self-locating prior on a local multiverse frame. For any possible multiverse size* n *such that* $1 < n < 1/p$ *(where* p *is the fine-tuning parameter), for any center* i *such that* $\Pr(E(i)) > 0$*,* n *universes are confirmed relative to one:*

$$\frac{\Pr_i(U_n)}{\Pr_i(U_1)} > \frac{\Pr(U_n)}{\Pr(U_1)}$$

or else $\Pr(U_1) = 0$.

Proof. $\Pr_i(-)$ clearly satisfies Agents: since $E(i)$ only contains non-null centers, $\Pr(I_0 \mid E(i)) = 0$. Since $\Pr_i(-)$ is an evidential posterior, all of the conditions of theorem 1 hold. □

Similarly, theorem 2 implies the following.

Theorem 4. *Let* $\Pr(-)$ *be a separable, evidentially natural self-locating prior on a local multiverse frame. Let* i *be a center such that* $\Pr(E(i)) > 0$ *and* $\Pr_i(-)$ *is ordinary (definition 7). For any possible multiverse size* $n > 1$*,* n *universes are confirmed relative to one.*

References

Arntzenius, Frank, and Cian Dorr. 2017. "Self-Locating Priors and Cosmological Measures". In *The Philosophy of Cosmology*, ed. by Khalil Chamcham et al., 396–428. Cambridge: Cambridge University Press.

— . Manuscript. "What to Expect in an Infinite World".

Borges, Jorge Luis. 1941. "La Biblioteca de Babel". In *El Jardín de Senderos que Bifurcan*. Editorial Sur.

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

Bradley, Darren. 2011. "Confirmation in a Branching World: The Everett Interpretation and Sleeping Beauty". *British Journal for the Philosophy of Science* 62 (2): 323–342. doi:10.1093/bjps/axq013.

— . 2009. "Multiple Universes and Observation Selection Effects". *American Philosophical Quarterly* 46 (1): 72.

Briggs, R.A. 2010. "Putting a Value on Beauty". *In Tamar Szabo Gendler and John Hawthorne (Eds.), Oxford Studies in Epistemology, Volume 3. Oxford University Press*: 3–34.

Builes, David. 2020. "Time-Slice Rationality and Self-Locating Belief". *Philosophical Studies* 177 (10): 3033–3049. doi:10.1007/s11098-019-01358-1.

Cappelen, Herman, and Josh Dever. 2013. *The Inessential Indexical: On the Philosophical Insignificance of Perspective and the First Person*. Oxford University Press.

Carnap, Rudolf. 1950. *Logical Foundations of Probability*. Chicago]University of Chicago Press.

Carter, Brandon. 1983. "The anthropic principle and its implications for biological evolution". *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences* 310 (1512): 347–363.

Chalmers, David J. 2011. "Frege's Puzzle and the Objects of Credence". *Mind* 120 (479): 587–635.

Dorr, Cian. Manuscript. "A Challenge for Halfers".

Draper, Kai, Paul Draper, and Joel Pust. 2007. "Probabilistic Arguments for Multiple Universes". *Pacific Philosophical Quarterly* 88 (3): 288–307.

Elga, Adam. 2004. "Defeating Dr. Evil with Self-Locating Belief". *Philosophy and Phenomenological Research* 69 (2): 383–396. doi:10.1111/j.1933-1592.2004.tb00400.x.

— . 2000. “Self Locating Belief and the Sleeping Beauty Problem”. *Analysis* 60 (2): 143–147.

Fitelson, Branden. 1999. “The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity”. *Philosophy of Science* 66 (3): 378. doi:10.1086/392738.

Greaves, Hilary. 2004. “Understanding Deutsch’s Probability in a Deterministic Universe”. *Studies in History and Philosophy of Modern Physics* 35 (3): 423–456. doi:10.1016/j.shpsb.2004.04.006.

Guth, Alan H. 2007. “Eternal inflation and its implications”. *Journal of Physics A: Mathematical and Theoretical* 40, no. 25 (): 6811–6826. ISSN: 1751-8121. doi:10.1088/1751-8113/40/25/s25. http://dx.doi.org/10.1088/1751-8113/40/25/S25.

Hacking, Ian. 1987. “The Inverse Gambler’s Fallacy: The Argument From Design. The Anthropic Principle Applied to Wheeler Universes”. *Mind* 96 (383): 331–340.

Halpern, Joseph. 2004. “Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems”. In *Proceedings of the Twentieth Conference on Uncertainty in Ai*, ed. by Tamar Szabo Gendler and John Hawthorne, 111–142. Oxford University Press.

Hawthorne, John. 2002. “Deeply Contingent A Priori Knowledge”. *Philosophy and Phenomenological Research* 65 (2): 247–269. doi:10.1111/j.1933-1592.2002.tb00201.x.

Hawthorne, John, and Yoaav Isaacs. 2018. “Fine-Tuning Fine-Tuning”. In *Knowledge, Belief, and God: New Insights in Religious Epistemology*, ed. by Matthew A. Benton, John Hawthorne, and Dani Rabinowitz, 136–168. Oxford: Oxford University Press.

Hawthorne, John, and Maria Lasonen-Aarnio. 2009. “Knowledge and Objective Chance”. In *Williamson on Knowledge*, ed. by Patrick Greenough and Duncan Pritchard, 92–108. Oxford University Press.

Holder, Rodney D. 2002. “Fine-tuning, Multiple Universes and Theism”. *Noûs* 36 (2): 295–312.

Jeffrey, Richard C. 1983. “Bayesianism With A Human Face”. In *Testing Scientific Theories*, ed. by John Earman, 133–156. University of Minnesota Press.

Juhl, Cory. 2006. “Fine-Tuning is Not Surprising”. *Analysis* 66 (4): 269–275. doi:10.1093/analys/66.4.269.

Keynes, John Maynard. 1921. *A Treatise on Probability*. Dover Publications.

Landsman, Klaas. 2016. "The Fine-Tuning Argument: Exploring the Improbability of Our Existence". In *The challenge of chance*, 111–129. Springer, Cham.

Leeds, Stephen. 2007. "Juhl on Many Worlds". *Noûs* 41 (3): 536–549. doi:10.1111/j.1468-0068.2007.00660.x.

Leslie, John. 1988. "No Inverse Gambler's Fallacy in Cosmology". *Mind* 97 (386): 269–272.

— . 1990. "Risking the world's end". *Interchange* 21 (1): 49–58.

— . 1989. *Universes*. Routledge.

Lewis, David. 1979. "Attitudes De Dicto and De Se". *Philosophical Review* 88 (4): 513–543.

Lewis, Geraint F., and Luke A. Barnes. 2017. *A Fortunate Universe: Life in a Finely-Tuned Cosmos*. Cambridge: Cambridge University Press.

Magidor, Ofra. 2015. "The Myth of the De Se". *Philosophical Perspectives* 29 (1): 249–283. doi:10.1111/phpe.12065.

Manley, David. Manuscript. "On Being a Random Sample".

McGrath, P. J. 1988. "The Inverse Gambler's Fallacy and Cosmology a Reply to Hacking". *Mind* 97 (386): 265–268.

McGrew, Timothy, Lydia McGrew, and Eric Vestrup. 2001. "Probabilities and the Fine-Tuning Argument: A Sceptical View". *Mind* 110 (440): 1027–1038. doi:10.1093/mind/110.440.1027.

Meacham, Christopher J. G. 2008. "Sleeping Beauty and the Dynamics of De Se Beliefs". *Philosophical Studies* 138 (2): 245–269.

— . 2016. "Ur-Priors, Conditionalization, and Ur-Prior Conditionalization". *Ergo: An Open Access Journal of Philosophy* 3. doi:10.3998/ergo.12405314.0003.017.

Moss, Sarah. 2012. "Updating as Communication". *Philosophy and Phenomenological Research* 85 (2): 225–248. doi:10.1111/j.1933-1592.2011.00572.x.

Ninan, Dilip. 2013. "Self-Location and Other-Location". *Philosophy and Phenomenological Research* 87 (1): 301–331. doi:10.1111/phpr.12051.

Olum, Ken D. 2004. "Conflict Between Anthropic Reasoning and Observation". *Analysis* 64 (1): 1–8. doi:10.1111/j.0003-2638.2004.00452.x.

— . 2002. “The Doomsday Argument and the Number of Possible Observers”. *Philosophical Quarterly* 52 (207): 164–184. doi:10.1111/1467-9213.00260.

Page, Don N. 1999. “Can quantum cosmology give observational consequences of many-worlds quantum theory?” *Eighth Canadian conference on general relativity and relativistic astrophysics*. doi:10.1063/1.1301589. http://dx.doi.org/10.1063/1.1301589.

Rees, Martin. 2008. *Just six numbers: The deep forces that shape the universe*. Basic Books.

Rota, Michael. 2005. “Multiple Universes and the Fine-Tuning Argument: A Response to Rodney Holder”. *Pacific Philosophical Quarterly* 86 (4): 556–576. doi:10.1111/j.1468-0114.2005.00242.x.

Russell, Jeffrey Sanford. 2020. “On the Probability of Plenitude”. *Journal of Philosophy* 117 (5): 267–292. doi:10.5840/jphil2020117517.

Salow, Bernhard. 2018. “The Externalist’s Guide to Fishing for Compliments”. *Mind* 127 (507): 691–728. doi:10.1093/mind/fzw029.

Schwarz, Wolfgang. 2017. “Diachronic Norms for Self-Locating Beliefs”. *Ergo: An Open Access Journal of Philosophy* 4. doi:http://dx.doi.org/10.3998/ergo.12405314.0004.025.

Sebens, Charles T., and Sean M. Carroll. 2016. “Self-Locating Uncertainty and the Origin of Probability in Everettian Quantum Mechanics”. *British Journal for the Philosophy of Science*, no. 1: axw004. doi:10.1093/bjps/axw004.

Stalnaker, Robert C. 2008. *Our Knowledge of the Internal World*. Oxford University Press.

Titelbaum, Michael. 2008. “The Relevance of Self Locating Beliefs”. *Philosophical Review* 117 (4): 555–606.

Titelbaum, Michael G. 2012. *Quitting Certainties: A Bayesian Framework Modeling Degrees of Belief*. Oxford University Press.

— . 2016. “Self-Locating Credences”. In *The Oxford Handbook of Probability and Philosophy*, ed. by Alan Hajek Christopher Hitchcock. Oxford University Press.

— . 2013. “Ten Reasons to Care About the Sleeping Beauty Problem”. *Philosophy Compass* 8 (11): 1003–1017. doi:10.1111/phc3.12080.

Vilenkin, Alexander. 1995. "Predictions from Quantum Cosmology". *Phys. Rev. Lett.* 74 (6): 846–849. doi:10.1103/PhysRevLett.74.846. https://link.aps.org/doi/10.1103/PhysRevLett.74.846.

— . 2011. *The Principle of Mediocrity*. arXiv: 1108.4990 [hep-th].

Weinberg, Steven. 1989. "The Cosmological Constant Problem". *Reviews of Modern Physics* 61 (1): 1.

— . 2000. "The Cosmological constant problems". In *4th International Symposium on Sources and Detection of Dark Matter in the Universe (DM 2000)*. arXiv: astro-ph/0005265.

Weintraub, Ruth. 2004. "Sleeping Beauty: A Simple Solution". *Analysis* 64 (1): 8–10. doi:10.1111/j.0003-2638.2004.00453.x.

Whitaker, M. A. B. 1988. "On Hacking's Criticism of the Wheeler Anthropic Principle". *Mind* 97 (386): 259–264.

White, Roger. 2000. "Fine-Tuning and Multiple Universes". *Noûs* 34 (2): 260–276. doi:10.1111/0029-4624.00210. Reprinted with a postscript in Neil A. Manson, ed. "Fine-Tuning and Multiple Universes". Chap. 13 in *God and Design: The Teleological Argument and Modern Science*, 229–250. London ; New York: Routledge, 2003.

— . 2003. "Fine-Tuning and Multiple Universes". Chap. 13 in *God and Design: The Teleological Argument and Modern Science*, ed. by Neil A. Manson, 229–250. London ; New York: Routledge.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford University Press.

Wilson, Alastair. 2013. "Everettian Confirmation and Sleeping Beauty". *British Journal for the Philosophy of Science*, no. 3: axt018. doi:10.1093/bjps/axt018.

Yli-Vakkuri, Juhani, and John Hawthorne. 2018. *Narrow Content*. Oxford University Press.