Centering Cosmological Calculations

Isaac Wilhelm

Penultimate draft of Jan 20, 2026.

Please cite the version in K. Kraay and D. Rubio (Eds.), *The Blackwell Companion to Philosophy and the Multiverse*.

Abstract

According to many cosmological theories, in extremely large universes—arguably, including ours—the probability of the current observational data, occurring somewhere or other, is one. To address confirmation problems which this generates, cosmologists have developed what are called 'first-person probabilities': roughly, probabilities of our seeing the observed data, rather than probabilities of the observed data occurring somewhere or other. This paper investigates the assumptions used to calculate first-person probabilities in cosmology. As I argue, certain justifications for those assumptions face problems stemming from exactly how first-person phenomena relate to third-person phenomena. These problems highlight the importance of exploring the foundational underpinnings of first-person probabilities in cosmological theorizing. And ultimately, these problems represent serious challenges for any attempt to reduce first-person probabilities to third-person probabilities.

1 Introduction

Various cosmological theories imply that in sufficiently large universes, it is virtually guaranteed that the observed data—about the cosmic microwave background spectrum, cosmological constant's value, and so on—will occur somewhere or other. In fact, it is virtually guaranteed that this data is duplicated, multiple times, throughout the universe. According to many such theories, the observed data occurs with probability around one.[1]

Therefore, in sufficiently large universes, the probability of thus-and-so data occurring cannot be used to distinguish substantively different cosmological theories from one another. For these different theories all predict that this probability equals one. And observations of trivial, unit probability data cannot be used to confirm some of these theories and disconfirm others.

In response to this, some have proposed a different approach to probability in cosmological theorizing. The approach introduces first-person probabilities, and distinguishes them from third-person probabilities (Srednicki & Hartle, 2010, p. 1; 2013, pp. 2-3).[2] Very roughly, first-person probabilities are assigned to what *we* observe, to *our* data – to information expressed using first-person, indexical locutions like 'I', 'we', 'our', and 'us'. And very roughly, third-person probabilities are assigned to data occurring somewhere or other in the universe – to information expressed using third-person, non-indexical locutions like names and definite descriptions. The probabilities mentioned earlier, of thus-and-so data occurring somewhere or other, are third-person probabilities – these are what cannot be used to distinguish substantively different cosmological theories. But that leaves open whether first-person probabilities can be used to distinguish substantively different cosmological theories. Perhaps first-person probabilities, if defined and calculated appropriately, could be used to confirm some cosmological theories over others. And indeed, many proposed definitions and calculations, of first-person probabilities in cosmology, seek to do just that (Azhar, 2015; Hartle & Hertog, 2017a; Hartle & Srednicki, 2007; Srednicki & Hartle, 2010).

This paper explores several cosmological approaches to first-person probabilities in large universes featuring many duplicates of our data. The approaches combine (i) assumptions about third-person probabilities, of the sort which cos-

[1]For instance, see theories of false vacuum eternal inflation (Guth, 1981; Vilenkin, 1983), chaotic inflation (Linde, 1983, 1986), and the landscape of string theory vacuum states (Freivogel et al., 2006; Kachru et al., 2003; Susskind, 2007).

[2]These are also occasionally called 'top-down probabilities' and 'bottom-up probabilities' respectively (Hawking & Hertog, 2006, p. 4; Hartle & Hertog, 2009, p. 1).

mological theories generate, with (ii) assumptions about first-person probabilities, concerning what we are likely to observe. These assumptions facilitate calculations of first-person probabilities which can, in principle at least, distinguish—for the purposes of confirmation and disconfirmation—different theories of cosmology. And in this paper, I investigate these assumptions' logical and conceptual foundations: the extent to which they are justified; and how to understand them.

One of the main philosophical motivations, for this paper, concerns the fact that cosmological theories have proposed reductions of first-person probabilities to third-person probabilities. The philosophical literature contains many arguments for the view that the first-personal does not, in general, reduce to the third-personal (Lewis, 1974, 1979; Ninan, 2010; Perry, 1979; Wilhelm, in press). Standard calculations of first-person probabilities in cosmology, however, challenge that: the calculations purport to reduce first-person probabilities to third-person probabilities. This paper explores that purported reductions in detail, to see whether or not they succeed; ultimately, I argue, they do not. So this paper uses cosmological probabilities to present a case study of the difficulties which arise when trying to reduce first-person, indexical phenomena to third-person, non-indexical phenomena.

Another motivation, for this paper, concerns the connection between the cosmological and philosophy literatures. The distinction between first-person probabilities and third-person probabilities, in cosmology, parallels the distinction between centered credence and uncentered credence in formal epistemology (Elga, 2000; Lewis, 2001; Wilhelm, 2021), the distinction between centered chance and uncentered chance in epistemology and philosophy of science (Wilhelm, 2022a, 2023), and the distinction between centered propositions and uncentered propositions in metaphysics and philosophy of language and philosophy of mind (Lewis, 1998; Ninan, 2013). Very roughly, centered propositions, also called 'first-person propositions' in what follows, are propositions that indexicals must be used to express; and uncentered propositions, some of which are called 'third-person propositions' in what follows, are propositions that indexicals need not be used to express. Centered credences and uncentered credences are agents' degrees of confidence assigned to centered propositions and uncentered propositions, respectively; and centered chances and uncentered chances are objective, worldly chances associated, respectively, with centered propositions and uncentered propositions. There is hardly any cross-disciplinary discussion, between cosmology and philosophy, of how the theories of first-person, centered phenomena in one field, might bear on the theories of first-person, centered phenomena in the other field. This paper demonstrates the benefits of more cross-disciplinary discussion: for example, as I show, philosophical theories of centered phenomena illuminate tensions between different assumptions regarding first-person probability in cosmology; and cosmological theories of centered phenomena illuminate precise, concrete strategies for calculating first-person probabilities that philosophy would do well to explore.[3]

In Section 2, I lay the groundwork for what follows. Specifically, I present the simple cosmological model on which I focus here, I summarize a calculation of third-person probabilities and the problems they face, I outline the general strategy for calculating first-person probabilities in cosmology, and I discuss the common assumption that certain probability distributions—used to calculate first-person probabilities—are uniform. In Section 3, I explore one calculation of first-person probabilities. This calculation assumes, roughly, that we are equally likely to be any of the observers, in the universe, who see the same data that we ourselves see. In Section 4, I explore another calculation of first-person probabilities. This calculation assumes, roughly, that we are equally likely to be any of the observers, in the universe, at all – regardless of the data which those observers see.

[3]One challenge, in holding cross-disciplinary discussions, concerns the different vocabulary that the different disciplines use, and the different ways in which those disciplines use that vocabulary. Some differences in vocabulary are merely verbal: they do not track importantly different assumptions about first-person, centered phenomena. Other differences in vocabulary are substantive: they do indeed track importantly different assumptions. And often, it is hard to tell whether a given difference in vocabulary is verbal or substantive: one important instance of this, discussed in Section 3, concerns the fact that cosmological theorizing formally represents the first-personality of first-person probabilities by superscripting the probability functions themselves with the symbol ‘$(1p)$’, whereas formal epistemological theorizing formally represents the first-personality of first-person probabilities by invoking centered propositions in the relevant probability functions’ arguments. Because this paper focuses on first-person theorizing in cosmology, I largely follow the vocabulary developed in the cosmological literature.

2 Preliminaries

In Section 2.1, I present a simple cosmological model which provides a basis for all later calculations of probabilities. In Section 2.2, I explain how to calculate third-person probabilities in the simple model, and I present four problems which those third-person probabilities face. In Section 2.3, I outline a general strategy for calculating first-person probabilities in the model. In Section 2.4, I discuss the plausibility—and in some cases, implausibility—of assuming uniform distributions in these calculations.

2.1 A Simple Model

The simple model is a universe containing N cycles over time. Let T be any theory which implies that each cycle is either red or blue. Think of the property of being red, and the property of being blue, as two different possible values for the cosmic microwave background temperature (Srednicki & Hartle, 2010, p. 3), or two different values of the cosmological constant (Hartle & Hertog, 2017a, p. 188). Think of each cycle as its own relatively isolated, mini-universe.

Let E be the proposition that at least one cycle contains observers, and let R be the proposition that at least one cycle is red. Let E, R[4] be the proposition that there exists a red cycle which contains an observer.[5] Suppose that E, R is precisely the data which we have: our evidence, in other words, is all and only that (i) at least one observer exists, and (ii) that observer sees a red cycle. Let $N_R(T)$ be the number of red cycles, according to theory T. And for each cycle, let p_E be the probability of there existing an observer in that cycle: so every cycle is as likely as every other cycle to contain an observer.

[4]In what follows, for all X and Y, 'X, Y' is roughly shorthand for '$(X \wedge Y)$'.

[5]Given that E has the logical form of something like $\exists x Ex$ and R has the logical form of something like $\exists x Rx$, it seems that E, R has the logical form of something like $\exists x Ex \wedge \exists x Rx$ rather than—as the description of E, R in the main text stipulates—the logical form of something like $\exists x(Ex \wedge Rx)$. But this subtlety will not matter much here. So following the cosmological literature, I generally describe E and R and E, R as in the main text.

2.2 Calculating Third-Person Probabilities

Now to calculate third-person probabilities in the model just defined. For each cycle, $(1 - p_E)$ is the probability of that cycle not containing an observer. Therefore, since there are $N_R(T)$ red cycles, $(1 - p_E)^{N_R(T)}$ is the probability that no red cycles contain observers. And therefore, the probability that at least one red cycle contains an observer is $1 - (1 - p)^{N_R(T)}$. In other words,

$$P(E, R \mid T) = 1 - (1 - p_E)^{N_R(T)} \tag{1}$$

where the conditionalization on T appears because all of the probabilities in question presuppose that T is the correct theory of red and blue in the universe. This is a third-person probability.

In four ways, equation (1) is somewhat problematic for the purposes of scientific reasoning. First, in practice, (1) does not distinguish between different theories that posit lots of red cycles (Srednicki & Hartle, 2010, p. 3). To see why, consider two theories T_1 and T_2 such that both $N_R(T_1)$ and $N_R(T_2)$ are extremely large. For instance, perhaps T_1 is the theory that every cycle is red, and T_2 is the theory that every other cycle is red. Because of how big $N_R(T_1)$ and $N_R(T_2)$ are, equation (1) implies that $P(E, R \mid T_1) \approx 1$ and $P(E, R \mid T_2) \approx 1$. By Bayes' theorem, it follows that $P(T_1 \mid E, R) \approx \frac{1}{2}$ and $P(T_2 \mid E, R) \approx \frac{1}{2}$.[6] Therefore, (1) does not allow our data to distinguish between these two theories: given our data, we are more-or-less equally well-justified in believing that every cycle is red as in believing that around half of the cycles are red.

Second, if p_E is sufficiently close to 1, then (1) does not distinguish between many different theories (Srednicki & Hartle, 2007, p. 4). For if $p_E \approx 1$, then for any theories T_1 and T_2 which posit at least one red cycle, $P(E, R \mid T_1) \approx 1$ and $P(E, R \mid T_2) \approx 1$. So again, Bayes' theorem implies that $P(T_1 \mid E, R) \approx \frac{1}{2}$ and $P(T_2 \mid E, R) \approx \frac{1}{2}$. And so again, for large p_E, (1) implies that given our data, we

[6] In particular, assuming that for all theories T distinct from T_1 and T_2, $P(T) = 0$, and assuming that $P(T_1) = P(T_2)$, Bayes' theorem implies that $P(T_1 \mid E, R) = \frac{P(E,R|T_1)P(T_1)}{\sum_i P(E,R|T_i)P(T_i)} = \frac{P(E,R|T_1)P(T_1)}{P(E,R|T_1)P(T_1)+P(E,R|T_2)P(T_2)} = \frac{P(E,R|T_1)}{P(E,R|T_1)+P(E,R|T_2)} \approx \frac{1}{1+1} = \frac{1}{2}$. $P(T_2 \mid E, R) \approx \frac{1}{2}$ follows by an analogous argument.

are more-or-less equally well-justified in believing many different theories.

Third, a more conceptual issue: (1) does not have the right content for the purposes of *our* theorizing. For (1) does not specify the first-person likelihoods of what *we* observers might see. (1) specifies a third-person likelihood: the likelihood, conditional on T, of at least one observer existing and seeing red. But what matters, for the purposes of *our* scientific inquiry, is *us*. For our purposes, what matters is not how some observers or other, in a red cycle, make predictions and formulate explanations and manage expectations and just generally navigate the world. What matters is how *we ourselves* do all that.

Since this has implications for the approaches to first-person likelihoods discussed later, it is worth describing the issue in more detail. Towards that end, let W_E be the proposition that we exist, let W_R be the proposition that we see red, and let W_E, W_R be the proposition that we exist and see red. Then W_E, W_R and E, R are numerically distinct propositions. They are related, of course: W_E, W_R logically entails E, R, since if we observers exist in a red cycle, then as a matter of logic, at least one observer exists in a red cycle. But E, R does not entail W_E, W_R. For given the fact that at least one observer exists in a red cycle, it does not follow that we exist in a red cycle; just consider a possibility in which the extant observer is not us. Put another way, no law of nature guarantees that in every physical possibility, we automatically 'get to be' whatever observer—in that physical possibility—exists in a red cycle; no law guarantees that so long as there exists at least one observer in a red cycle, that observer must be us.[7] So W_E, W_R and E, R are not even true in all the same possibilities. Forget hyperintensional equivalence (Berto & Nolan, 2023): they are not even intensionally equivalent (Stalnaker, 1987).

Hence the third problematic feature of (1): since E, R and W_E, W_R are different propositions, it follows that specifying probabilities for E, R – namely, for the proposition that there exists at least one observer in a red cycle – is different from specifying probabilities for W_E, W_R – namely, for the proposition that we

[7]Proposing such a law would, in fact, risk inconsistency. To see why, suppose the candidate law were this: so long as there exists at least one observer in a red cycle, that observer is us. Then consider a possibility in which there exist two distinct observers in red cycles. By the candidate law, we are identical to both observers. By the transitivity of identity, the distinct observers are identical; contradiction.

exist in a red cycle. Again, (1) specifies a third-person likelihood, conditional on T, of E, R. But for the purposes of *our* scientific inquiry, about *our* place in the universe—and given that there are many different observers which we might be—that is somewhat beside the point. What matters is the first-person likelihood, conditional on T, of W_E, W_R: the first-person likelihood, conditional on T, that we observers exist and see red.

Fourth, and relatedly: (1) does not provide a helpful guide for observers embedded within the physical system that theory T describes; that is, within the universe as a whole. The third-person likelihoods in (1) might facilitate making predictions, and setting expectations, for observers outside of the system: observers with a god-like view of the universe, say. But in certain situations, third-person likelihoods do not facilitate making certain predictions, and setting certain expectations, for observers within the system being studied. This happens when the system contains duplicates of the observers at issue. For then the observes cannot tell—on the basis of the third-person theory T alone—which duplicate they are. And for the purposes of making certain sorts of predictions, and setting certain sorts of expectations, that matters.

An example will illustrate why. Let o_1 and o_2 be duplicate observers, exactly one of whom is us. Suppose that o_1 and o_2 know all the same third-person facts. So for example, perhaps they both are in a red cycle, so they both know the third-person fact—namely E, R—that at least one red cycle contains an observer; and perhaps they both calculated the third-person likelihood of E, R conditional on T. Then both o_1 and o_2 can use their third-person facts to make predictions, and set their expectations, regarding what o_1 will see and what o_2 will see. Unfortunately for us, none of that settles what *we* will see. The third-person facts, in our evidence, have no implications whatsoever regarding whether we are o_1 or o_2: our being o_1 is compatible with the third-person facts, and our being o_2 is compatible with the third-person facts. And this matters, because given different theories, o_1 and o_2 might eventually see very different things: one of them might be a Boltzmann brain which, because it is about to fluctuate out of existence, never sees anything else at all (Carroll, 2019).

The issue here is related to the fact that E, R and W_E, W_R are distinct propositions. We want to make a prediction for what we observers in a red cycle might see. But that is distinct from making a prediction about what at least one observer in a red cycle might see. The third-person probability given by (1) facilitates predictions for o_1 and for o_2. But that is distinct from facilitating predictions for us. What we need, of course, is a theory of how third-personal facts and probabilities, about o_1 and o_2, connect to first-personal facts and probabilities about us. And that connection is precisely what (1) fails to provide.

These four problems show that for the purposes of reasoning in a large universe that conforms to many contemporary cosmological theories, third-person probabilities like (1) are unideal. Those third-person probabilities fail to distinguish between importantly different theories. And those third-person probabilities do not specify the likelihoods for first-person propositions about what we observers, embedded in the universe, might see.

2.3 The General Strategy for Calculating First-Person Probabilities

First-person probabilities can be used to address all four problems. For first-person probabilities can, arguably, distinguish between importantly different theories. And first-person probabilities, if defined and calculated correctly, promise to specify the likelihoods for what we observers might see.

In order to calculate the relevant first-person probabilities, Azhar (2015), Hartle and Hertog (2017a), Srednicki and Hartle (2010), and others, follow a general strategy consisting of two steps. First, posit a probability function ξ—called a 'xerographic distribution'—which assigns first-person probabilities to propositions about whether we are this-or-that observer in the universe. Second, calculate the first-person probabilities for our observational data by combining (i) the posited distribution ξ, with (ii) the third-person probabilities that cosmological theories generate. Different xerographic distributions, and different combinations of that distribution with third-person probabilities, yield different first-person likelihoods for our data – and so in principle at least, yield different predictions which can then be tested against observations and experiments.

Here are more details about the general strategy. The first step posits a xerographic distribution ξ. Intuitively, ξ conveys the probability that we are this-or-that observer in the universe. Because of that, the probabilities expressed by ξ are first-personal: they convey the probability that we are this-or-that, rather than the probability that observer o_1, say, is this-or-that.

There are many different choices for ξ – many different ways, that is, of distributing probabilities over the observers which we might be. The following principle describes one intuitive schema for distributing probabilities.

> The Indifference$_L$ Principle
>
> Let L be a non-empty set of cycles, in the universe, which contain observers. Then for each $\ell \in L$, $\xi^L_\ell = \frac{1}{|L|}$ is the probability—relative to L—that we are the observers in ℓ.[8]

Different choices of L yield different xerographic distributions ξ^L. And different xerographic distributions correspond to different views about which observers, in the universe, are candidates for the observers that we might be.

Two choices for L are discussed most often in the literature: the non-empty set of cycles containing any observers whatsoever, and the non-empty set of cycles containing observers whose data is precisely our data E, R (Azhar, 2015, p. 5; Srednicki & Hartle, 2010, pp. 4-5). In what follows, let '*typO*' denote the former set and let '*typD*' denote the latter set. They are called '*typO*' and '*typD*' because the Indifference$_L$ Principle is often interpreted as asserting that we are typical among the observers at issue. The xerographic distribution ξ^{typO} says that we are typical among all observers whatsoever – so we are equally likely to be any one of those observers – and the xerographic distribution ξ^{typD} says that we are typical among all observers with our data E, R – so we are equally likely to be any one of the observers with that data.[9]

The second step in the general strategy combines a xerographic distribution with third-person probabilities, in order to calculate probabilities that are first-

[8] For versions of this principle, see (Azhar, 2015, p. 4; Hartle & Hertog, 2017a, p. 202; Srednicki & Hartle, 2010, p. 2).

[9] This description of ξ^{typO} and ξ^{typD} identifies certain probabilistic distributions with typicality facts. According to the theories of typicality that I prefer, there are important differences between typicality and probability: being typical should not always be equated with being equiprobable, for instance (Wilhelm, 2022b). But this is probably a merely verbal difference, and so will not matter here.

personal. This involves making lots of choices about what various first-personal probabilities, conditional on various facts about ξ and the theory T and other propositions, are. Many formal, conceptual, physical, and philosophical assumptions underlie those choices. This paper conducts a detailed investigation of these assumptions for particular implementations of the general strategy: one implementation, based on ξ^{typD}, is discussed in Section 3; and another implementation, based on ξ^{typO}, is discussed in Section 4.

An aside: in what follows, for brevity and concreteness, I focus on implementations of the general strategy developed by Srednicki and Hartle (2010). My remarks apply to other implementations of the general strategy, for instance the generalizations developed by Azhar (2015) or the applications discussed by Hartle and Hertog (2017a). For lack of space, however, I do not discuss that here.

2.4 Uniform Distributions

The two particular implementations of the general strategy, discussed later, posit uniform probability distributions. For instance, each employs one of the xerographic distributions ξ^{typO} and ξ^{typD} – and by the Indifference$_L$ Principle, each of those distributions is uniform over the sets of cycles *typO* and *typD* respectively. Moreover, other posited distributions are uniform too: for instance, in Section 4, I posit several uniform distributions to replace a uniform distribution which Srednicki and Hartle posit.

The philosophical literature contains various arguments for positing uniform distributions in cases of uncertainty over duplicate situations (Builes, 2024; Elga, 2004; Eva, 2019; Liu, 2024; Pettigrew, 2016; Williamson, 2018). For example, suppose I am uncertain about exactly which duplicate, of N different duplicates, I am – and suppose that these duplicates all have the same total evidence. Then according to the conclusions of these arguments, for each duplicate, my credence in the first-person proposition that I am that duplicate should be $\frac{1}{N}$.

Though I am sympathetic to positing uniform distributions—as proponents of the general strategy for calculating first-person probabilities tend to be—my reasons differ from the reasons that these philosophical arguments provide. I doubt that uniform distributions can be justified by purely, or even primarily, a priori arguments. What justifies a posited uniform distribution, in a given cosmological model, is simply that model's empirical success. If a posited theory T and posited xerographic distribution ξ^L make an incorrect prediction, then either T or ξ^L must be rejected. If uniform distributions have purely a priori justification, then T must be rejected and ξ^L—being uniform and so a priori—must be retained. But that is implausible, and in fact, borders on being unscientific. Like any posit in any scientific theory, ξ^L is open to revision: dogmatically adhering to ξ^L, come what may, runs against scientific practice. So contrary to what some philosophical arguments might suggest, rejecting the uniform distribution ξ^L should not be ruled out a priori.

Proponents of the general strategy agree with this. As Hartle and Hertog note, there is not a shred of observational evidence for positing uniform xerographic distributions (2017b, p. 12). Such posits are simple and natural, of course: but they should be subjected to experimental testing, as any scientific assumption should be. As Srednicki and Hartle argue, in situations involving Boltzmann brains, we should not adopt a uniform first-person xerographic distribution over observers with our evidence (2013, p. 3). That would generate problematic predictions.

Besides, there are well-known cases in which non-uniform distributions, over duplicate agents with the same total evidence, are preferable to uniform distributions. Take the Everett interpretation of quantum mechanics. According to standard versions of the Everett interpretation, at the present time t, there are many exact physical duplicates of me, you, and our community as a whole.[10] But given the empirical frequencies in our total evidence, it makes no sense to posit a uniform probability distribution over these duplicates: it would be irrational to think that I am just as likely to be any one duplicate as any other. For the empirical frequencies support a non-uniform probability distribution: the $|\psi|^2$ distribution given by the Born rule. So the rational credences to adopt, given the empirical frequencies in my total evidence, are those which match the non-uniform $|\psi|^2$ distribution (Wallace, 2012).

[10]In fact, there may well be continuously many exact physical duplicates of all that: see (Deutsch, 1985, p. 20; Wilhelm, 2023, p. 318).

This is a strong argument against the claim that uniform probability distributions, over duplicate observers with the same total evidence, are justified a priori. Perhaps it is plausible, or reasonable, to posit a uniform probability distribution—over duplicate observers with the same total evidence—in specific, restricted cases. But those posits are defeasible, just as any posit in any scientific theory is defeasible. And in fact, Everettian quantum mechanics provides a straightforward empirical counterexample to those sorts of posits.

So understand the discussion of uniform distributions, in what follows, along empiricist lines. The xerographic literature focuses on uniform distributions, and so I do too – and that is fine. But do not interpret this as endorsing the a priority of, say, the Indifference$_L$ Principle. Interpret this in accord with what Srednicki and Hartle (2010, p. 3), and many others, say: uniform xerographic distributions provide simple and natural cases to study, but we should be open to other, non-uniform distributions over observers.

3 Calculations Based on TypD

In this section, I explore an implementation of the general strategy based on the xerographic distribution ξ^{typD}: the implementation uses ξ^{typD} to offer a reduction of first-person probabilities to third-person probabilities. As I argue, however, the implementation faces several problems. Some problems, which Srednicki and Hartle themselves discuss, concern the fact that the calculated first-person probabilities—like the third-person probabilities from Section 2.2—cannot be used to distinguish importantly different theories. Other problems, which I discuss at length, stem from the fact that these first-person probabilities fail to distinguish between propositions like E and propositions like W_E: between the proposition that at least one observer exists, that is, and the proposition that we exist. Still other problems concern the fact that important assumptions, used to calculate these first-person probabilities, are in tension with one another: the justification for one assumption, in particular, supports denying another assumption. Ultimately, these problems suggest that the reduction of first-person probabilities to third-person probabilities has not been successful.

To begin the calculation, take $L = typD$ in the Indifference$_L$ Principle. So for all $\ell \in typD$, $\xi_\ell^{typD} = \frac{1}{|typD|}$. Therefore, given any cycle ℓ containing observers with E, R as their data, $\frac{1}{|typD|}$ is the probability that we are the observers in ℓ.

Srednicki and Hartle use $P^{(1p)}(R \mid E, T, \xi^{typD})$ to formally represent an important first-person probability about us seeing red (2010, p. 4). In particular, their formalism endorses the principle below.

The "We See Red" Representation

$P^{(1p)}(R \mid E, T, \xi^{typD})$ is the proper way to formally represent the first-person probability that we see red, conditional on E and T and ξ^{typD}.

In other words, the first-person probability that we see red—conditional on at least one observer existing, the theory, and our being in one of the cycles in *typD* with equal probability—just is the probability expressed by the formal term $P^{(1p)}(R \mid E, T, \xi^{typD})$.

There are reasons to doubt the "We See Red" Representation. R was defined as the proposition that there exists at least one red cycle. As discussed earlier, that is different from the proposition W_R that we see red. The two propositions are related: the proposition that we see red implies the proposition that at least one red cycle exists; W_R, that is, implies R. But they are distinct, for reasons analogous to the reasons—presented while discussing the third and fourth problems from Section 2.2—why W_E, W_R and E, R are distinct, that is. Given that at least one red cycle exists, it does not follow that we see red: for we could have been one of the observers in a blue cycle. So contrary to what the "We See Red" Representation claims, $P^{(1p)}(R \mid E, T, \xi^{typD})$ is not the proper way to formally represent the first-person probability that we see red.

There is another way to understand what $P^{(1p)}(R \mid E, T, \xi^{typD})$ represents, however. Just endorse the principle below.

The Second "We See Red" Representation

$P^{(1p)}(R \mid E, T, \xi^{typD})$ is the proper way to formally represent the first-person probability that at least one red cycle exists, conditional on E and T and

ξ^{typD}.

So $P^{(1p)}(R \mid E, T, \xi^{typD})$ is best understood as assigning a probability to the third-person proposition R that at least one red cycle exists, conditional on—among other things—the first-person proposition that we are observers in one of the red cycles which ξ^{typD} assigns a probability of $\frac{1}{|typD|}$.

To continue: Srednicki and Hartle claim that since ξ^{typD} is nonzero only on cycles with observers that have E, R as their data, the following must hold.

$$1 = P^{(1p)}(R \mid E, T, \xi^{typD}) \tag{2}$$

And this seems plausible. For the first-person probability that at least one red cycle exists, conditional on—among other things—that we are observers in a red cycle, is obviously 1.[11] That is a perfectly good justification of (2).

Note that conditionalizing on ξ^{typD}, so understood, amounts to conditionalizing on the first-person proposition that we are observers in a red cycle which ξ^{typD} assigns a probability of $\frac{1}{|typD|}$ – or more succinctly, on the first-person proposition that we are observers in a red cycle. That is a reasonable way of understanding conditionalization on ξ^{typD} because that is, indeed, the main conditional information that ξ^{typD} conveys. In addition, that is how Srednicki and Hartle describe conditionalizing on ξ^{typD}: they informally describe $P^{(1p)}(E \mid T, \xi^{typD})$ as the first-person probability that we exist given that we are in a red cycle (2010, p. 4). For future purposes, it will be helpful to have a name for this view: so call this the 'We Are In Red' interpretation of conditionalization on ξ^{typD}. And so the We Are In Red interpretation, of conditionalization on ξ^{typD}, is what justifies (2).

One more qualification is worth flagging: it concerns the difference between (i) how philosophers tend to formally represent first-person phenomena, and (ii) how $P^{(1p)}(R \mid E, T, \xi^{typD})$ represents first-person phenomena. Philosophers employ a formalism that locates first-personality in the propositions being assigned probabilities, rather than in the probability functions which do the assigning. So

[11] Note that in addition to this just being obvious, it also follows from probability theory: since W_R implies R, the probability axioms imply that $P^{(1p)}(W_R \mid E, T, \xi^{typD}) \leqslant P^{(1p)}(R \mid E, T, \xi^{typD})$; and because $P^{(1p)}(W_R \mid E, T, \xi^{typD})$ must be 1—since conditional on us being in a *typD* cycle, it is guaranteed that we see red—(2) follows.

philosophers would denote the first-person probability that we see red as $P(W_R)$: no superscript on the P is needed, since the first-personality of this expression is encoded in the proposition W_R. In the expression $P^{(1p)}(R \mid E, T, \xi^{typD})$, conversely, first-personality is located in the probability function, not in the propositions assigned probabilities.[12] Hence the "We See Red" Representation, which uses a first-person probability function to 'reinterpret' claims like R – about at least one red cycle existing – into claims like W_R – about us seeing red.

There are interesting formal and conceptual questions about how these different approaches to locating first-personality in probabilistic expressions—in the propositions which serve as arguments for probability functions, or in the probability functions themselves—relate. Perhaps there are substantive differences between the sorts of first-personal phenomena which these different approaches can be used to capture. Or perhaps these different approaches are equivalent in some appropriate sense.

For my part, I suspect that (i) these different approaches are not equivalent, and (ii) locating first-personality in probability functions, rather than those functions' arguments, can obfuscate problems for the attempted reduction of first-person probabilities to third-person probabilities. For instance, it is hard to see how an equation like $P^{(1p)}(R) = \frac{1}{2}$ counts as expressing anything genuinely first-personal: it simply asserts that a third-person proposition – namely, that at least one red cycle exists – has a probability – which, apart from being described with the English epithet 'first-person', and apart from being bequeathed the superscript '$(1p)$', has nothing distinctively first-personal about it – of $\frac{1}{2}$. It is easier to see how an equation like $P(W_R) = \frac{1}{2}$ counts as expressing something first-personal: it asserts that a first-person proposition – namely, that we see red, which is different from the third-person proposition that at least one red cycle exists – has a probability – understood perhaps as the rational credence of some agent, or as an objective chance – of $\frac{1}{2}$. And to date at least, no one has proposed a translation scheme that maps third-person propositions X to first-person propositions $X^{(1p)}$

[12] This description of the difference is a bit quick: for in $P^{(1p)}(R \mid E, T, \xi^{typD})$, first-personality appears both in the probability function $P^{(1p)}$ and also in the understanding of conditionalization on ξ^{typD} discussed earlier. But for the purposes of explaining the difference between philosophers' formal representation of first-person phenomena, and the formal representation that Srednicki and Hartle employ, that does not matter.

in a way which, among other things, homomorphically preserves compositional structure and logical entailment and other important semantic properties of the propositions that are assigned probabilities. And most importantly, no one has proposed a truth-preserving or probability-preserving translation between claims like $P^{(1p)}(X) = x$ and corresponding claims like $P(X^{(1p)}) = x$.

But regardless, my main point is this. Srednicki, Hartle, and other physicists, have developed an approach to first-personality which focuses on probability functions rather than the propositions serving as those functions' arguments. Formal epistemologists, metaphysicians, and philosophers of mind, have developed approaches to first-personality which focus on those functions' arguments—the propositions, or the sentences featuring indexicals—rather than the functions themselves. It is worth exploring, in more detail, how these approaches relate.

Back to the calculation of first-person probabilities based on ξ^{typD}. Given equation (2), and standard facts about conditional probabilities, it follows that

$$\begin{aligned}
1 &= P^{(1p)}(R \mid E, T, \xi^{typD}) \\
&= \frac{P^{(1p)}(R, E, T, \xi^{typD})}{P^{(1p)}(E, T, \xi^{typD})} \\
&= \frac{P^{(1p)}(R, E, T, \xi^{typD})}{P^{(1p)}(T, \xi^{typD})} \frac{P^{(1p)}(T, \xi^{typD})}{P^{(1p)}(E, T, \xi^{typD})} \\
&= \frac{P^{(1p)}(E, R \mid T, \xi^{typD})}{P^{(1p)}(E \mid T, \xi^{typD})}
\end{aligned}$$

Therefore,

$$P^{(1p)}(E, R \mid T, \xi^{typD}) = P^{(1p)}(E \mid T, \xi^{typD}) \qquad (3)$$

In other words, the first-person probability that at least one observer exists and sees red, conditional on T and ξ^{typD}, equals the first-person probability that at least one observer exists, conditional on T and ξ^{typD}.

Next, Srednicki and Hartle claim that the first-person probability that we exist, given that we are in a red cycle—in other words, the first-person probability of E conditional on, among other things, ξ^{typD}—equals the third-person probability that at least one observer exists in a red cycle (2010, p. 4). Put formally,

$$P^{(1p)}(E \mid T, \xi^{typD}) = P(E, R \mid T) \tag{4}$$

Note that this equation offers an identification of a first-person probability with a third-person probability. So in that sense, this equation reduces the first-person probability of E – conditional on T and ξ^{typD} – to the third-person probability of E, R – conditional on T.

Finally, equations (1) and (4) jointly imply that $P^{(1p)}(E \mid T, \xi^{typD}) = 1 - (1 - p_E)^{N_R(T)}$. So by (3),

$$P^{(1p)}(E, R \mid T, \xi^{typD}) = 1 - (1 - p_E)^{N_R(T)} \tag{5}$$

Equation (5) completes this particular reduction of first-person probabilities to third-person probabilities. For according to (5), the first-person probability that we exist in a red cycle is just the third-person probability that at least one observer-occupied red cycle exists.

For several reasons, however, equation (5) is problematic. It implies that first-person probabilities are no more able to distinguish between theories—or more accurately, between pairings of different theories with the posited xerographic distribution ξ^{typD}—than third-person probabilities are (Srednicki & Hartle, 2010, p. 4). In particular, first-person probabilities—as calculated here using the xerographic distribution ξ^{typD}—face the first and second problems discussed in Section 2.2. They cannot distinguish between different theories that posit lots of red cycles; and if p_E is sufficiently close to 1, then they cannot distinguish between lots of other different theories too.

Equation (5) is problematic for other, more conceptual reasons as well. The reasons concern equation (4), which asserts that the first-person probability that we exist, given that we are in a red cycle, is the same as the third-person probability that at least one observer exists in a red cycle. For two reasons, this is questionable. The first reason is basically a version of the third and fourth problems discussed in Section 2.2: the proposition that we exist, W_E, is not the same as the proposition that at least one observer exists, E. Again, as one illustration of the difference, note that while W_E entails E, it is not the case that E entails W_E. So $P^{(1p)}(E \mid T, \xi^{typD})$ is not best understood as assigning a first-person probability to us existing. Rather, $P^{(1p)}(E \mid T, \xi^{typD})$ is best understood as assigning a first-person probability to the third-person proposition that, conditional on T and ξ^{typD}, at least one observer exists. Better to represent the first-person probability of us existing, conditional on T and ξ^{typD}, as $P^{(1p)}(W_E \mid T, \xi^{typD})$.

Here is another, stronger way to put the point. As explained in Section 2, given the fact that at least one observer exists, it does not follow that we exist. For again, just consider a possibility in which the extant observer is not us. No law of nature guarantees that in every physical possibility, we automatically 'get to be' whatever observer—in that physical possibility—exists there. No law guarantees that so long as there exists at least one observer, that observer must be us. So no law of nature implies that $P^{(1p)}(E \mid T, \xi^{typD})$ can serve as a stand-in for $P^{(1p)}(W_E \mid T, \xi^{typD})$.

The problem can be illuminated by considering it from the perspective of alien observers, elsewhere in the universe. Suppose the proposition that we exist, conditional on our being in a red cycle, just is the proposition that at least one observer exists in a red cycle – as the problematic interpretation of equation (4) suggests. Any reason we have, to endorse this, is equally well a reason for alien observers to endorse something analogous: the proposition that they exist, conditional on their being in a red cycle, just is the proposition that at least one observer exists in a red cycle. But then by the transitivity of identity, the proposition that we exist, conditional on our being in a red cycle, just is the proposition that they exist, conditional on their being in a red cycle. And that is wrong: propositions about us existing are not equivalent to propositions about those alien observers existing.

The second reason why (4) is problematic: regardless of whether $P^{(1p)}(E \mid T, \xi^{typD})$ is understood as assigning a first-person probability to us existing or as assigning a first-person probability to at least one observer existing, $P^{(1p)}(E \mid T, \xi^{typD})$ should not be identified with $P(E, R \mid T)$. The reason: given the We Are In Red interpretation of conditionalization on ξ^{typD}, it follows that $P^{(1p)}(E \mid T, \xi^{typD})$ is 1. For conditional on us being observers in a red cycle, the first-person probability of us existing and also the first-person probability of at least one observer existing – that is, the first-person probability of W_E and also the first-person probability of E – must be 1: for if we observers exist, then trivially, the fact that we exist follows; and if we observers exist, then trivially, the fact that at least one observer exists follows. In other words, regardless of how $P^{(1p)}(E \mid T, \xi^{typD})$ is interpreted, $P^{(1p)}(E \mid T, \xi^{typD}) = 1$; so either way, $P^{(1p)}(E \mid T, \xi^{typD})$ is not identical to $P(E, R \mid T)$; and so either way, (4) does not hold. Because versions of this issue will crop up later, it is worth giving it a name: so call this the 'trivialization consequence of conditionalizing on xerographic distributions'.

The problem here stems from tension between the justification for (2) and the justification for (4). Recall that in justifying (2), Srednicki and Hartle make the reasonable claim that conditional on ξ^{typD}, the first-person probability that we see red, or the first-person probability that at least one red cycle exists, is 1: for conditional on us being observers in a red cycle – that is, conditional on ξ^{typD} – it follows trivially that 1 is the first-person probability of at least one red cycle existing, and it follows trivially that 1 is the probability of our cycle being red. But it is just as reasonable to claim that conditional on ξ^{typD}, the first-person probability that we exist, or the first-person probability that at least one observer exists, is 1 too: for conditional on us being observers in a red cycle – that is, conditional on ξ^{typD} – it follows trivially that 1 is the first-person probability of at least one observer existing, and it follows trivially that 1 is the probability of us existing. In other words, by replacing 'red cycle' with 'observer' and 'our cycle being red' with 'us existing', the justification for (2) extends to justify

$$1 = P^{(1p)}(E \mid T, \xi^{typD}) \tag{6}$$

rather than (4). So (5) does not hold: the first-person probability that we exist in a red cycle is not the third-person probability that at least one red cycle exists.

One might respond by denying the We Are In Red interpretation. The occurrence of 'ξ^{typD}' in '$P^{(1p)}(E \mid T, \xi^{typD})$', one might claim, merely indicates that the first-person probability is informed, in some way or other, by a xerographic distribution which assigns $\frac{1}{|typD|}$ to every cycle in $typD$ and assigns 0 to every other cycle. So for instance, instead of representing that first-person probability by the notation '$P^{(1p)}(E \mid T, \xi^{typD})$', it would have been equally reasonable to represent that first-person probability by the notation '$P^{(1p)}_{\xi^{typD}}(E \mid T)$' or something like that. Then one could claim that $P^{(1p)}_{\xi^{typD}}(E \mid T) = P(E, R \mid T)$, in accord with (4). And that, one might conclude, avoids the trivialization consequence of conditionalizing on xerographic distributions.

There are two different problems with this objection. The first problem: this objection does not actually say how the first-person probability at issue—represented notationally by either '$P^{(1p)}(E \mid T, \xi^{typD})$' or '$P^{(1p)}_{\xi^{typD}}(E \mid T)$'—is informed by the xerographic distribution ξ^{typD}. How does (i) the fact that we are one of the observers in $typD$ with equal probability, determine (ii) the fact that the first-person probability of E, conditional on T, equals the third-person probability of E, R conditional on T? The objection has no answer. So the objection has not shown how xerographic distributions are relevant to reducing first-person probabilities to third-person probabilities.

The second problem: this objection undermines the justification for (2). The justification, recall, went like this: given the We Are In Red interpretation of conditionalization on ξ^{typD}, it follows trivially that the first-person probability of R, conditional on ξ^{typD}, is 1. But if the We Are In Red interpretation of conditionalization on ξ^{typD} does not hold, then we no longer have reason to endorse the equation $1 = P^{(1p)}(R \mid E, T, \xi^{typD})$. The first-person probability at issue, represented notationally by '$P^{(1p)}(R \mid E, T, \xi^{typD})$', could be represented notationally by '$P^{(1p)}_{\xi^{typD}}(R \mid E, T)$' instead. But it is not at all obvious that this first-person probability is 1. In fact, that seems false: the existence of at least one red cycle, given that at least one observer exists and given T, is not guaranteed. So this objection cannot be used to rescue the present calculation of first-person probabilities from the second problem.

Note that given all this, there is still a way to derive a first-person probability for E, R. For (2) and (6) jointly imply

$$1 = P^{(1p)}(E, R \mid T, \xi^{typD}) \tag{7}$$

In other words, conditional on ξ^{typD} – that is, conditional on us existing in a red cycle – the first-person probability of E, R – understood either as the first-person probability that we exist in a red cycle, or as the first-person probability that at least one observer exists in a red cycle – is 1.

This result, though useless from a practical point of view, is completely unsurprising, and indeed intuitively correct. Given the interpretation of the first-person probability of E, R at issue, (7) makes a lot of sense. For conditional on the fact that we observers exist in a red cycle—that is, conditional on the fact that we are one of the *typD* observers with probability $\frac{1}{|typD|}$—the existence of at least one observer in a red cycle is guaranteed. Hence (7).

There are two upshots of all this. The first upshot is in agreement with Srednicki and Hartle's big-picture conclusion. Conditionalizing on ξ^{typD}, in calculations of the relevant first-person probabilities, does indeed trivialize the results of those calculations.

The second upshot disagrees with Srednicki and Hartle, however, over that in which the trivialization consists. Whereas Srednicki and Hartle take the trivialization to consist in the relevant first-person probabilities reducing to third-person probabilities that do not distinguish between different cosmological theories, the above arguments show that the trivialization consists in the relevant first-person probabilities all equalling 1. The reason why they all equal 1: conditional on ξ^{typD}, it follows that 1 is the probability of us existing, of us being in a red cycle, of at least one observer existing, of at least one observer being in a red cycle, and of all combinations thereof. So there is no reduction of first-person probabilities to third-person probabilities. Instead, there is just a plausible and unsurprising, but also uninformative, result about what the relevant first-person probabilities are: equation (7).

4 Calculations Based on TypO

Because of these issues, it is worth looking at another implementation of the general strategy for calculating first-person probabilities. Hence this section, which explore an implementation that invokes the xerographic distribution ξ^{typO} – recall that *typO* is the non-empty set of cycles containing observers, and ξ^{typO} is the xerographic distribution which assigns each cycle a probability of $\frac{1}{|typO|}$. As I argue, the first-person probabilities generated by this implementation face many of the problems which the first-person probabilities from Section 3 face:

specifically, they do not reduce to third-person probabilities. In addition, however, the first-person probabilities generated by this implementation also face problems connected to the proper way to understand conditionalizing on ξ^{typO} specifically: basically, there are two ways to understand this conditionalization, each of which generates different first-person probabilities, and each of which is appropriate for a different range of situations.

To begin the calculation, take $L = typO$ in the Indifference$_L$ Principle. So for all $\ell \in typO$, $\xi_\ell^{typO} = \frac{1}{|typO|}$. Therefore, given any cycle ℓ containing observers, $\frac{1}{|typO|}$ is the probability that the cycle ℓ observers are us.

As before, the goal is to calculate the first-person probability of our data: in this case, $P^{(1p)}(E, R \mid T, \xi^{typO})$. And to start, as before, Srednicki and Hartle assume a version of the "We See Red" Representation: the first-person probability that we see red—conditional on E, T, and our being an observer in a *typO* cycle with equal probability $\frac{1}{|typO|}$—is represented by $P^{(1p)}(R \mid E, T, \xi^{typO})$ (2010, p. 2). This incurs all the problems of the "We See Red" Representation discussed in Section 3: R is different from W_R, so $P^{(1p)}(R \mid E, T, \xi^{typO})$ cannot be interpreted as assigning a probability to the first-person proposition that we see red, conditional on this-and-that. For now, however, set this aside.

Srednicki and Hartle present two different calculations of first-person probabilities based on ξ^{typO}: one based on symmetries of their model (2010, p. 4), and another based on combinatorics (2010, pp. 7-8). For lack of space, I focus on the former; the same basic points apply to the latter as well.

An important aside: these calculations yield first-person probabilities that can distinguish between pairings of different cosmological theories with the xerographic distribution ξ^{typO}. So the resulting first-person probabilities avoid the first and second problems discussed in Section 2.2: the problem of failing to distinguish between theories that posit lots of red cycles, for instance. Therefore, in what follows, I focus on the extent to which the resulting first-person probabilities avoid other, more conceptual problems – like problems stemming from the attempt to reduce these first-person probabilities to something third-personal.

There are two key equations which the calculation assumes. The first key equation, stated below, is left implicit.

$$P^{(1p)}(E \mid T, \xi^{typO}) = P(E \mid T) \tag{8}$$

Note that this equation offers an identification of a first-person probability with a third-person probability. So in that sense, this equation reduces the first-person probability of E – conditional on T and ξ^{typO} – to the third-person probability of E – conditional on T.

This equation, being the analog of equation (4) from Section 3, faces two problems analogous to the problems which (4) faces. First, $P^{(1p)}(E \mid T, \xi^{typO})$ is arguably not best understood as assigning a first-person probability to us existing. Rather, since the proposition that we exist – namely, W_E – is distinct from the proposition that at least one observer exists – namely, E – $P^{(1p)}(E \mid T, \xi^{typO})$ is arguably best understood as assigning a first-person probability to the third-person proposition that at least one observer exists. Again, no law of nature guarantees that in every physical possibility, we automatically 'get to be' whatever observer exists there.

Second, regardless of whether $P^{(1p)}(E \mid T, \xi^{typO})$ is understood as assigning a first-person probability to us existing or as assigning a first-person probability to at least one observer existing, $P^{(1p)}(E \mid T, \xi^{typO})$ should not be identified with $P(E \mid T)$. For just as conditionalization on ξ^{typD} should be understood in accord with the We Are In Red interpretation, conditionalization on ξ^{typO} should be understood in accord with the following 'We Are Observers' interpretation: conditionalizing on ξ^{typO} amounts to conditionalizing on the first-person proposition that we are observers in a cycle which ξ^{typO} assigns a probability of $\frac{1}{|typO|}$; or more succinctly, on the first-person proposition that we are in one of the observer-occupied cycles. And given the We Are Observers interpretation of conditionalization on ξ^{typO}, it follows that $P^{(1p)}(E \mid T, \xi^{typO})$ must be 1. For conditional on us being one of the observers in a *typO* cycle, the existence of at least one observer follows automatically. This is a version of the trivialization consequence of conditionalizing on xerographic distributions.

As before, one might respond by denying the We Are Observers interpretation of conditionalization on ξ^{typO} – perhaps the relevant first-person probability could be equally well represented as '$P^{(1p)}_{\xi^{typO}}(E \mid T)$'. But as before, this objection faces problems. For instance, it remains unclear how the xerographic distribution ξ^{typO} contributes to determining the first-person probability at issue, the one represented by either '$P^{(1p)}(E \mid T, \xi^{typO})$' or '$P^{(1p)}_{\xi^{typO}}(E \mid T)$'.

One qualification is worth flagging, however. In what follows, I explore an alternative way of interpreting conditionalization on ξ^{typO}, that differs in some respects from the We Are Observers interpretation. But more on this alternative later.

The second key equation is below.

$$P^{(1p)}(R \mid E, T, \xi^{typO}) = \frac{N_R(T)}{N} \tag{9}$$

In other words, the first-person probability that at least one red cycle exists—conditional on at least one observer existing, T, and our being in one of the *typO* cycles with equal probability—is the fraction of red cycles in the universe.

In support of (9), Srednicki and Hartle provide the following argument (2010, p. 4). We are equally likely to exist in any cycle, since whether a cycle contains observers is probabilistically independent of whether that cycle is red or blue. Therefore, the probability that we see red—conditional on at least one observer existing, T, and our being in one of the *typO* cycles with equal probability—is the probability that our cycle is red. And the latter probability, of our cycle being red, is $\frac{N_R(T)}{N}$. Hence (9).

A quick aside: (9) does not face a version of the trivialization consequence of conditionalizing on xerographic distributions. For given the We Are Observers interpretation of conditionalization on ξ^{typO}, $P^{(1p)}(R \mid E, T, \xi^{typO})$ need not be 1. Given that we are in a *typO* cycle, it does not follow automatically that we see red; some observers in *typO* cycles might see blue.

Nevertheless, another potential problem—related to, but distinct from, the trivialization consequence of conditionalizing on xerographic distributions—arises for (9). The potential problem is that contrary to Srednicki and Hartle's argument in support of (9), it does not seem to be the case that conditional on our being in one of the specific *typO* cycles with equal probability, as well as conditional on E and T, we are equally likely to exist in any cycle of the universe whatsoever.

For conditional on our being in one of the specific *typO* cycles, the probability of our existing in a cycle outside of *typO* is 0. Therefore, given a specific, fixed choice of *typO*—of exactly which cycles contain observers—(9) does not provide the correct first-person probability for our seeing red conditional on, among other things, our being in one of the *typO* cycles.

Instead, plausibly, the correct first-person probability for our seeing red—conditional on, among other things, our being in one of the *typO* cycles—is given by the equation below, where $N_{R,O}$ is the total number of observers in red cycles.[13]

$$P^{(1p)}(R \mid E, T, \xi^{typO}) = \frac{N_{R,O}}{|typO|} \tag{10}$$

In other words, the first-probability of our seeing red—conditional on E, T, and the fact that we are one of the observers in a cycle belonging to the specific set *typO*—is the fraction of red cycles among *typO* cycles.

The above argument, against (9) and in favor of (10), was based on what can be called the 'Fixed $TypO$' assumption: *typO*, as invoked in $P^{(1p)}(R \mid E, T, \xi^{typO})$, is a specific, fixed, particular set of cycles. Given this assumption, it follows that conditionalizing on ξ^{typO} amounts to, among other things, conditionalizing on the information that those specific cycles exactly—out of the N cycles in the universe—contain observers. To appreciate what this means, note that it is importantly different from conditionalizing on the information that, say, $k = |typO|$ of the N cycles in the universe contain observers. For the latter is much less informative than the former: conditionalizing on (i) the fact that k out of the N total cycles contain observers, is much less informative than conditionalizing on (ii) the fact that *these k cycles specifically, out of the N cycles—the k cycles in the specified set* typO*—contain observers.* For (i) does not take a stand on exactly what the set *typO*, of observer-containing cycles, actually is; (ii), obviously, does.

The Fixed $TypO$ assumption is motivated by much of Srednicki and Hartle's formalism.[14] And given this assumption, (10) is better than (9). For given this

[13]Note that $N_{R,O}$ is determined both by (i) the theory T, which fixes which cycles are red, and (ii) the set *typO*, which specifies exactly which cycles contain observers.

[14]For instance, Srednicki and Hartle describe the superscript '*typD*' in ξ^{typD} as meaning "typical in the class with [our data]" (2010, p. 4), whichever specific class that might be. So similarly, the superscript '*typO*' in ξ^{typO} seems to mean: typical in the specific class of cycles

assumption, it is not true that conditional on our being in one of the *typO* cycles specifically, we are equally likely to exist in any cycle of the universe whatsoever: the likelihood of our existing in a cycle outside *typO*, conditional on our being in a *typO* cycle, is 0. Rather, conditional on our being in one of the *typO* cycles specifically, the likelihood of our cycle being red is the fraction of red cycles in *typO*, as (10) implies.

Interestingly, a different assumption about how to understand *typO*, as invoked in $P^{(1p)}(R \mid E, T, \xi^{typO})$, supports (9) over (10). This assumption—call it the 'Variable *TypO*' assumption, is that *typO*, as invoked in $P^{(1p)}(R \mid E, T, \xi^{typO})$, is a variable. In other words, *typO* is not a specific, fixed, particular set of cycles. Rather, understand conditionalization on ξ^{typO} as conditionalizing only on the information that *typO*—the set of cycles containing observers—is some subset or other, of the set of universe cycles, with the specific, fixed, particular size k (and on the information that the probability of our being in any given one of those cycles is $\frac{1}{k}$); and think of k as the size which every candidate choice of *typO*, every way of assigning $|typO| = k$ observers to cycles in the universe, has in common. So whereas the Fixed *TypO* assumption treats the instance of '*typO*' in '$P^{(1p)}(R \mid E, T, \xi^{typO})$' like a name for a specific set, the Variable *TypO* assumption treats the instance of '*typO*' in '$P^{(1p)}(R \mid E, T, \xi^{typO})$' like a variable which ranges over all sets of a certain specific size. And the Variable *TypO* assumption supports the following alternative to the We Are Observers interpretation of conditionalization on ξ^{typO}: according to the alternative, conditionalization on ξ^{typO} amounts conditionalizing only on the information that (i) the set of cycles containing observers has size $k = |typO|$, and (ii) our cycle is in that set, whatever that set might be. Call this the 'We Are Variable Observers' interpretation of conditionalization on ξ^{typO}.

The Variable *TypO* assumption, and the We Are Variable Observers interpretation, have a lot in their favor. They are suggested by the calculation of first-person probabilities in Srednicki and Hartle's Appendix B (2010, pp. 7-8), for instance.[15] And when combined with other plausible assumptions, they even

which contain observers.

[15] To see why, note that in the calculation, the first-person probability of our data conditional on ξ^{typO} and T is identified with the product of a xerographic distribution and a third-person probability summed over all choices of (i) the number of observer-occupied cycles $n_O \leqslant N$, (ii)

imply (9).

To see why, let $k \leqslant N$ represent the number of cycles containing observers, and let $Red \subseteq \{1, \ldots, N\}$ represent the set of red cycles. Assume that each distribution of the k observers, among the N cycles, is just as likely to be the true distribution of observers as each other. More formally, let $\{1, \ldots, N\}$ represent the N cycles of the universe, and assume that for any $typO \subseteq \{1, \ldots, N\}$ such that $|typO| = k$,

$$P(typO) = \frac{1}{\binom{N}{k}} \tag{11}$$

where of course, $\binom{N}{k}$ is the number of different ways of choosing k cycles from the N cycles of the universe; this is the assumption which most directly presupposes the Variable $TypO$ assumption. In addition, assume that the probability of any given cycle being red – that is, being in Red – conditional on that cycle being one which contains observers – that is, being in some given $typO$ – is just the fraction of red cycles among the $typO$ cycles. More formally, assume that for every $typO \subseteq \{1, \ldots, N\}$ such that $|typO| = k$,

$$P(Red \mid typO) = \frac{|typO \cap Red|}{|typO|} \tag{12}$$

This assumption is analogous to (10): for this, like (10), says that the probability of a given cycle being red, conditional on that cycle being in the specific set of cycles $typO$, is the fraction of red cycles among the $typO$ cycles. Also, assume that Red and $typO$ are probabilistically independent of E, T. More formally, assume

the number of red cycles among those observer-occupied cycles $n_R \leqslant n_O$, and (iii) which of those red, observer-occupied cycles is ours. Formally, this is written as follows: $P^{(1p)}(E, R \mid \xi^{typO}, T) = \sum_{n_O=1}^{N} \sum_{n_R=1}^{n_O} \sum_{\ell=1}^{n_R} \xi_\ell^{typO} P(n_O, n_R \mid T)$, where ξ_ℓ^{typO} is the probability that we are in cycle ℓ of $typO$ and $P(n_O, n_R \mid T)$ is the third-person probability of n_O cycles containing observers and n_R of those cycles being red, conditional on T; it is assumed that the probability of our being in cycle ℓ, of any given $typO$, is independent of the probability of n_O being the number of observer-occupied cycles and n_R being the number of those cycles which are red, conditional on T. The triple sum is basically taken over different choices for $typO$ – including over different candidate sizes k. So this calculation basically generalizes the Variable $TypO$ assumption from (i) concerning a single choice of k specifically, to (ii) all candidate choices of k whatsoever.

that for every $typO \subseteq \{1, \ldots, N\}$ such that $|typO| = k$,

$$\begin{aligned} P(Red \mid E, T, typO) &= P(Red \mid typO) \\ P(typO \mid E, T) &= P(typO) \end{aligned} \tag{13}$$

Finally, assume that the first-person probability of us seeing red, conditional on E and T and ξ^{typO}, is just the third-person probability of *Red*, conditional on E and T. More formally, assume that for every $typO \subseteq \{1, \ldots, N\}$ such that $|typO| = k$,

$$P^{(1p)}(R \mid E, T, \xi^{typO}) = P(Red \mid E, T) \tag{14}$$

where this is basically just a definition of what is meant by a first-person probability conditionalized on ξ^{typO}. Then

$$\begin{aligned} P^{(1p)}(R \mid E, T, \xi^{typO}) &= P(Red \mid E, T) \\ &= \sum_{\substack{typO \subseteq \{1,\ldots,N\} \\ |typO| = k}} P(typO \mid E, T) P(Red \mid E, T, typO) \\ &= \sum_{\substack{typO \subseteq \{1,\ldots,N\} \\ |typO| = k}} P(typO) P(Red \mid typO) \\ &= \sum_{\substack{typO \subseteq \{1,\ldots,N\} \\ |typO| = k}} \frac{1}{\begin{pmatrix} N \\ k \end{pmatrix}} \frac{|typO \cap Red|}{|typO|} \\ &= \frac{1}{k \begin{pmatrix} N \\ k \end{pmatrix}} \sum_{\substack{typO \subseteq \{1,\ldots,N\} \\ |typO| = k}} |typO \cap Red| \\ &= \frac{1}{N} \frac{(k-1)!\big((N-1)-(k-1)\big)!}{(N-1)!} \sum_{\substack{typO \subseteq \{1,\ldots,N\} \\ |typO| = k}} |typO \cap Red| \\ &= \frac{1}{N} \frac{1}{\begin{pmatrix} N-1 \\ k-1 \end{pmatrix}} \sum_{\substack{typO \subseteq \{1,\ldots,N\} \\ |typO| = k}} |typO \cap Red| \\ &= \frac{|Red|}{N} \end{aligned}$$

$$= \frac{N_R(T)}{N}$$

where the first equation is (14), the second equation follows from the law of total probability, the third equation follows from (13), the fourth equation follows from (11) and (12), the eighth equation follows from standard combinatorial identities, and the ninth equation follows from the notational fact that $|Red| = N_R(T)$. So given several reasonable assumptions, including the Variable $TypO$ assumption and the We Are Variable Observers interpretation of conditionalization on ξ^{typO}, (9) follows.

To summarize, here is the dialectical situation. There are two ways of understanding what it is to conditionalize on the information which a xerographic distribution contains. According to the We Are Observers interpretation, supported by the Fixed $TypO$ assumption, conditionalizing on ξ^{typO} amounts to conditionalizing on a specific xerographic distribution, which makes a specific choice about exactly which cycles have observers and exactly which cycles do not. This amounts to conditionalizing on a specific choice about exactly what the subset $typO$ is. According to the We Are Variable Observers interpretation, supported by the Variable $TypO$ assumption, conditionalizing on ξ^{typO} amounts to conditionalizing on a range of different xerographic distributions, which agree with one another insofar as they are all uniform over the same total number of cycles, but which disagree with one another insofar as they correspond to different choices of the specific set of cycles over which they are uniform. This amounts to conditionalizing over all the different choices of the subset $typO$—subject to the condition that those different choices have the same number of cycles in them—each of which is just as probable as each other.

Neither of these ways of understanding conditionalization on ξ^{typO} is better, full-stop, than the other. Certain definitions and notations, which Srednicki and Hartle employ, suggest the Fixed $TypO$ assumption – and therefore, support (10) over (9). But ultimately, no single understanding of conditionalization on ξ^{typO} is correct for all situations whatsoever. The correct understanding—and therefore, which of (10) or (9) holds—varies from situation to situation.

For instance, in many realistic situations, we do not know exactly which cycles contain observers. And in situations like that, the Variable $TypO$ assumption, coupled with the We Are Variable Observers interpretation, is more appropriate than the Fixed $TypO$ assumption coupled with the We Are Observers interpretation. For in situations like that, we do not know exactly which subset of the set of N cycles contains all and only the observers. Perhaps we do know, however, that the number of observer-occupied cycles is some specific $k \leqslant N$. So when calculating the first-person probability of R, we should conditionalize only on the relatively uninformative fact that *typO* is one of the k-sized subsets of the set of N cycles – that is, it makes perfect sense to endorse the We Are Variable Observers interpretation.

But in other realistic situations, for one reason or another, we know which specific cycles contain observers. And in situations like that, the Fixed $TypO$ assumption, coupled with the We Are Observers interpretation, is more appropriate than the Variable $TypO$ assumption coupled with the We Are Variable Observers interpretation. For in situations like that, we know a lot more than the relatively uninformative fact that for some specific $k \leqslant N$, *typO* is one of the k-sized subsets of the set of N cycles. We know the specific k-sized set *typO* of cycles containing observers. So when calculating the first-person probability of R, we conditionalize on the relatively informative fact that the set of cycles containing observers is that specific *typO* – that is, it makes perfect sense to endorse the We Are Observers interpretation.

So to be clear: the lesson here is not that (9) is wrong and must be replaced by (10). The lesson, rather, is that the difference between (9) and (10) tracks two different ways of understanding conditionalization on xerographic distributions. And each understanding is well-suited to a different situation, corresponding to what we might know about the distribution of observers through the universe.

Srednicki and Hartle use equations (8) and (9) to calculate first-person probabilities based on the xerographic distribution ξ^{typO} (2010, p. 4). Since

$$P(E \mid T) = \big(1 - (1 - p_E)^N\big) \tag{15}$$

equations (8) and (9) jointly imply that

$$P^{(1p)}(E, R \mid T, \xi^{typO}) = P^{(1p)}(R \mid E, T, \xi^{typO})P^{(1p)}(E \mid T, \xi^{typO})$$

$$
\begin{aligned}
&= \frac{N_R(T)}{N} P(E \mid T) \\
&= \frac{N_R(T)}{N} \left(1 - (1 - p_E)^N\right) \qquad (16)
\end{aligned}
$$

But as the above arguments show, (16) is not a good representation of the first-person probability that—conditional on T and on our being one of the *typO* observers with equal probability—we exist and see red. For equation (8) is problematic: it conflates the proposition that we exist and see red – namely W_E, W_R – with the proposition that at least one observer exists and sees red – namely, E, R – and in addition, given the conditionalization on ξ^{typO}, $P^{(1p)}(E \mid T, \xi^{typO})$ must be 1. And equation (9), while less problematic, is still limited: it only applies in situations where the Variable $TypO$ assumption holds, and so where the We Are Variable Observers interpretation of conditionalization on ξ^{typO} is best. So (16) does not successfully reduce the first-person probability of our seeing red, conditional on E and T and ξ^{typO}, to the third-personal probability corresponding to the product of (i) the fraction $\frac{N_R(T)}{N}$ of red cycles in the universe, and (ii) the third-person probability, expressed by equation (15), of at least one observer-occupied cycle existing.

5 Conclusion

To wrap up, here is a summary of the main problems arising for standard calculations of first-person probabilities in cosmology.

- Calculations Based on TypD
  1. Since the proposition that we see red differs from the proposition that at least one red cycle exists, $P^{(1p)}(R \mid E, T, \xi^{typD})$ is not best understood as assigning a first-person probability to the proposition that, conditional on E and T and ξ^{typD}, we see red. Rather, $P^{(1p)}(R \mid E, T, \xi^{typD})$ is best understood as assigning a first-person probability to the proposition that, conditional on E and T and ξ^{typD}, a red cycle exists.
  2. Similarly, since the proposition that we exist differs from the proposition that at least one observer exists, $P^{(1p)}(E \mid T, \xi^{typD})$ is not best understood

as assigning a first-person probability to the proposition that, conditional on T and ξ^{typD}, we exist. Rather, $P^{(1p)}(E \mid T, \xi^{typD})$ is best understood as assigning a first-person probability to the proposition that, conditional on T and ξ^{typD}, at least one observer exists.

3. $P^{(1p)}(E \mid T, \xi^{typD})$ does not equal $P(E, R \mid T)$. For given how other equalities are justified, conditionalization on ξ^{typD} should be understood as conditionalization on the claim that we are, with equal probability, one of the observers in a *typD* cycle. And conditional on that, E is guaranteed; so $P^{(1p)}(E \mid T, \xi^{typD})$ equals 1.

- Calculations Based on TypO
  1. For reasons analogous to those given earlier, $P^{(1p)}(R \mid E, T, \xi^{typO})$ is best understood as assigning a first-person probability to the proposition that, conditional on E and T and ξ^{typO}, a red cycle exists, rather than as assigning a first-person probability to the proposition that, conditional on E and T and ξ^{typO}, we see red.
  2. Similarly, $P^{(1p)}(E \mid T, \xi^{typO})$ is best understood as assigning a first-person probability to the proposition that, conditional on T and ξ^{typO}, at least one observer exists, rather than as assigning a first-person probability to the proposition that, conditional on T and ξ^{typO}, we exist.
  3. For reasons analogous to those given earlier, $P^{(1p)}(E \mid T, \xi^{typO})$ does not equal $P(E \mid T)$. For conditionalization on ξ^{typO} should be understood as conditionalization on the claim that we are, with equal probability, one of the observers in a *typO* cycle. And conditional on that, E is guaranteed; so $P^{(1p)}(E \mid T, \xi^{typO})$ equals 1.
  4. There is an important distinction between (i) conditionalizing first-person probabilities on a specific, fixed choice of the subset *typO* of cycles containing observers, and (ii) conditionalizing first-person probabilities on all the different candidate choices of the subset *typO*, where those different choices all contain the same specific, fixed number of cycles.
     – Suppose that conditionalization on ξ^{typO} is understood in accord with (i). Then the first-person probability of at least one red cycle existing—conditional on, among other things, ξ^{typO}—is given by $P^{(1p)}(R \mid E, T, \xi^{typO}) = \frac{N_{R,O}}{|typO|}$. This is most appropriate for situations in which we know exactly

which subset, of the set of N cycles in the universe, contains observers.

– Suppose that conditionalization on ξ^{typO} is understood in accord with (ii). Then the first-person probability of at least one red cycle existing—conditional on, among other things, ξ^{typO}—is given by $P^{(1p)}(R \mid E, T, \xi^{typO}) = \frac{N_R(T)}{N}$. This is most appropriate for situations in which we know exactly how many cycles, of the N cycles in the universe, contain observers – but not when we know exactly which cycles those are.

So several subtle issues arise for the proposed reductions of first-person probabilities to third-person probabilities in cosmology.

Ultimately, I am doubtful that any purported reduction of first-person probabilities to third-person probabilities will succeed. First-person facts about what I see, in the universe, are just too loosely connected to third-person facts about what thus-and-so observer sees, such-and-such observer sees, and so on. So I suspect that first-person probabilities, assigned to first-person facts about who I am and what I see, do not reduce to third-person probabilities assigned to third-person facts about what thus-and-so observer sees. Of course, these concerns might prove misguided: the reduction of first-person probabilities to third-person probabilities—and more generally, of first-personal phenomena to third-personal phenomena—remains an important, open research question, certainly worth pursuing. As I have argued here, however, any proposed reductions have significant problems to overcome.

Acknowledgements

Thanks to Emily Adlam, Jacob Barandes, Baptiste Le Bihan, Kenneth Boyce, Miles K. Donahue, Nikk Effingham, Thomas Hertog, Michael Townsen Hicks, Klaas Kraay, Han Li, Neil A. Manson, Christopher Meacham, John Pittard, Daniel Rubio, Brad Saad, Christopher Weaver, and the audience at the 2025 Multiverse and Philosophy Conference, for much helpful feedback and discussion.

References

Azhar, F. (2015). Testing typicality in multiverse cosmology. *Physical Review D*, *91*, 1–13.

Berto, F., & Nolan, D. (2023). *Hyperintensionality.* Retrieved from `https://plato.stanford.edu/entries/hyperintensionality/`

Builes, D. (2024). Eight arguments for first-person realism. *Philosophy Compass*, *e12959*, 1–14.

Carroll, S. M. (2019). Why boltzmann brains are bad. In S. Dasgupta, R. Dotan, & B. Weslake (Eds.), *Current controversies in philosophy of science* (pp. 7–20). New York, NY: Routledge.

Deutsch, D. (1985). Quantum theory as a universal physical theory. *International Journal of Theoretical Physics*, *24*(1), 1–41.

Elga, A. (2000). Self-locating belief and the sleeping beauty problem. *Analysis*, *60*(2), 143–147.

Elga, A. (2004). Defeating dr. evil with self-locating belief. *Philosophy and Phenomenological Research*, *69*(2), 383–396.

Eva, B. (2019). Principles of indifference. *The Journal of Philosophy*, *116*(7), 390–411.

Freivogel, B., Kleban, M., Martínez, M. R., & Susskind, L. (2006). Observational consequences of a landscape. *Journal of High Energy Physics*, *39*(3), 1–18.

Guth, A. H. (1981). Inflationary universe. *Physical Review D*, *23*(2), 347–356.

Hartle, J., & Hertog, T. (2009). Replication regulates volume weighting in quantum cosmology. *Physical Review D*, *80*(063531), 1–8.

Hartle, J., & Hertog, T. (2017a). The observer strikes back. In K. Chamcham, J. Silk, J. D. Barrow, & S. Saunders (Eds.), *The philosophy of cosmology* (pp. 181–205). Cambridge: Cambridge University Press.

Hartle, J., & Hertog, T. (2017b). One bubble to rule them all. *Physical Review D*, *95*(123502), 1–15.

Hartle, J., & Srednicki, M. (2007). Are we typical? *Physical Review D*, *75*(123523), 1–6.

Hawking, S., & Hertog, T. (2006). Populating the landscape. *Physical Review D*, *73*(123527), 1–9.

Kachru, S., Kallosh, R., Linde, A., & Trivedi, S. P. (2003). de Sitter vacua in string theory. *Physical Review D*, *68*, 1–10.

Lewis, D. (1974). Radical interpretation. *Synthese*, *23*, 331–344.

Lewis, D. (1979). Attitudes de dicto and de se. *The Philosophical Review*, *88*(4), 513–543.

Lewis, D. (1998). Index, context and content. In D. Lewis (Ed.), *Papers in philosophical logic* (pp. 21–44). Cambridge: Cambridge University Press.

Lewis, D. (2001). Sleeping beauty. *Analysis*, *61*(3), 171–176.

Linde, A. D. (1983). Chaotic inflation. *Physics Letters*, *129B*(3-4), 177–181.

Linde, A. D. (1986). Eternally existing self-reproducing chaotic inflationary universe. *Physics Letters B*, *175*(4), 395–400.

Liu, S. (2024). How to be indifferent. *Noûs*, 1–18.

Ninan, D. (2010). De se attitudes. *Philosophical Compass*, *5*(7), 551–567.

Ninan, D. (2013). Self-location and other-location. *Philosophy and Phenomenological Research*, *87*(2), 301–331.

Perry, J. (1979). The problem of the essential indexical. *Noûs*, *13*(1), 3–21.

Pettigrew, R. (2016). Accuracy, risk, and the principle of indifference. *Philosophy and Phenomenological Research*, *92*(1), 35–59.

Srednicki, M., & Hartle, J. (2007). Are we typical? *Physical Review D*, *75*, 123523.

Srednicki, M., & Hartle, J. (2010). Science in a very large universe. *Physical Review D*, *81*, 123524.

Srednicki, M., & Hartle, J. (2013). The xerographic distribution. *Journal of Physics*, *462*, 012050.

Stalnaker, R. (1987). *Inquiry.* Cambridge, MA: MIT Press.

Susskind, L. (2007). The anthropic landscape of string theory. In B. Carr (Ed.), *Universe or multiverse?* (pp. 247–266). Cambridge: Cambridge University Press.

Vilenkin, A. (1983). Birth of inflationary universes. *Physical Review D*, *27*(12), 2848–2855.

Wallace, D. (2012). *The emergent multiverse.* New York, NY: Oxford University Press.

Wilhelm, I. (2021). Centering the Principal Principle. *Philosophical Studies*, *178*(3), 1897–1915.

Wilhelm, I. (2022a). Centering the Everett Interpretation. *The Philosophical Quarterly*, *72*(4), 1019–1039.

Wilhelm, I. (2022b). Typical. *The British Journal for the Philosophy of Science*, *73*, 561–581.

Wilhelm, I. (2023). Centering the Born Rule. *Quantum Reports*, *5*, 311–324.

Wilhelm, I. (in press). Centering update. *The Journal of Philosophy*.

Williamson, J. (2018). Justifying the principle of indifference. *European Journal for Philosophy of Science*, *8*, 559–586.