Empirical Confirmation in Multiverse Cosmology

A Senior Thesis submitted to

the Glynn Family Honors Program

and the Department of Philosophy,

University of Notre Dame

Completed under the supervision of

Dr. Feraz Azhar

Signature of Supervisor

Enosh Paul Kallely Sholey

April 2025

Empirical Confirmation in Multiverse Cosmology

Enosh Paul Kallely Sholey

April 2025

Abstract

This paper is concerned with a specific issue in the philosophy of cosmology that arises in the context of inflationary models that describe our universe as one among many in a larger 'multiverse'. In the particular idealized setting of interest, all universes are causally disconnected from each other, so we have observational evidence only about our universe. In this context, asking how we should interpret evidence to conduct empirical confirmation for multiverse theories is an interesting question. Some, like Vilenkin and Bostrom, have argued that we ought to treat our evidence as if we are randomly selected from a reference class of observers in the multiverse and expect our evidence to correspond to the evidence a typical observer sees. This assumption is known as the typicality assumption, and Vilenkin and Bostrom have argued that we should prefer multiverse theories that describe us as typical observers over those that do not. Others, like Srednicki, Hartle, and Adlam have criticized this position and argued that we cannot assume we are typical. In this paper, I analyze the arguments given by proponents of typicality and assess whether they are justified in the multiverse context. I conclude that they are not justified in this context and furthermore, there is a way to conduct empirical confirmation for multiverse theories that does not make use of any assumptions about typicality or atypicality.

1 Introduction

The theory of cosmological inflation was developed in the early 1980s by Guth (1981), Linde (1982) and several others (Albrecht and Steinhardt, 1982), to combat the issue of the flatness problem, horizon problem, and the monopole problem, which plagued an otherwise successful Big Bang cosmology. Since the introduction of the theory, it has garnered support in contemporary cosmology owing to its ability to resolve these problems and also explain the anisotropies present in the Cosmic Microwave Background. However, one peculiar aspect of inflation continues to generate controversy in cosmology and the philosophy of cosmology. This aspect has to do with most inflationary models describing a so-called 'multiverse', where our universe is one of a large (potentially infinite) number of bubble universes within a larger multiverse.[1] To avoid confusion, the term 'universe' will be used to mean a single bubble universe while the term 'multiverse' will denote the collection of all bubble universes. There are three features of the multiverse setting that are important for the discussions that follow:

- In this idealized multiverse setting, all universes are causally disconnected from each other. This is a reasonable idealization as mechanism of inflation is such that the bubble universes are receding away from each other at speeds faster than the speed of light. This means that, if one were to send a signal at light speed from one universe to its nearest neighbor, it would never reach its destination, as the neighboring universe itself is moving away faster than the signal speed. Since nothing from one universe can ever reach another one, they are causally isolated from one another.

- Universes could look very different from each other. It has been postulated (Linde, 2017, Susskind, 2003) that the values of fundamental constants (such as the cosmological constant) in physics could change from universe to universe. These variations could produce universes that are wildly different from one another. For example, it could be that while our universe is dominated by matter, other universes are dominated by anti-matter.

- We do not know which bubble universe we are in. There are multiple universes that are consistent with our observational data. If we were in any one of these universes, our observations

[1] Bubble universes are also frequently called 'pocket universes', however for the purpose of clarity we will refer to bubble universes only.

would appear the same. Therefore, based solely on our data, we cannot determine which particular universe is ours.

To illustrate the debate this paper is concerned with, consider a toy model of the multiverse presented in figure 1. For this toy model, we will assume the multiverse is finite and contains only 6 bubble universes. Each bubble universe is indexed as U1, U2, U3 etc.[2] Consider also a toy cosmological theory which says that most universes (U1, U2, U3, U4) in the multiverse are antimatter dominated while the remaining few (U5, U6) are matter dominated. Assume also that we have a theory of human evolution that describes each universe as containing a single human.[3] One way to test such a cosmological theory would be to extract predictions about what we should observe given the theory and check if we really see what the theory predicts. Here we encounter a problem, namely that there is some ambiguity in what exactly the theory says we should see. What we observe about our universe will depend on which universe we are in, but we do not know which universe we are in. The only evidence we have is that our universe is matter dominated. How can we use this information for empirical confirmation of this theory if we don't know which universe we are in? To see the weight of the issue, consider a rival theory which says that U1, U2, U3 and U4 are matter dominated while the remaining two are antimatter dominated. How should we assign credences for these two theories? If we knew we were in a particular universe such as U6, then we can argue that since one theory says U6 is matter dominated and the other says U6 is antimatter dominated, we can prefer the theory that correctly describes U6 (our universe) as matter dominated. But we do not know which universe is our universe and herein lies the ambiguity between the theory and what the theory describes we should see.[4]

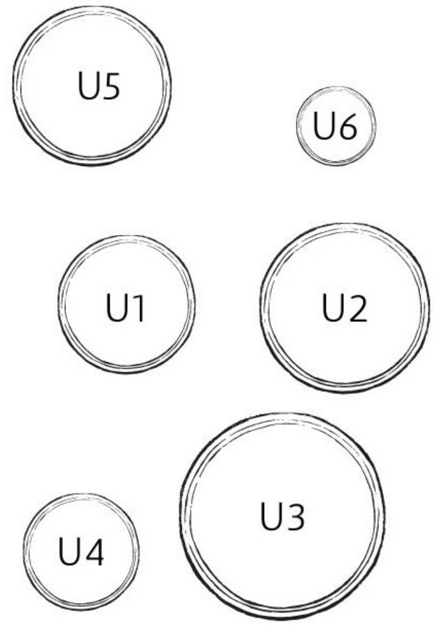

Figure 1

[2]In practice, a significant number of inflationary models describe the multiverse as inflating eternally into the future so that the number of bubble universes are infinite. Even in models with finite inflation into the future, the number of bubble universes tend to be extremely large. For simplicity we assume a multiverse with only 6 bubble universes in this toy model.

[3]This is of course, a simple toy model. We force only one human in a universe to keep the typicality considerations simple.

[4]There are two important topics that are in the background to discussions about typicality, namely the anthropic principle (Carter, 1974) and the measure problem. The anthropic principle may be roughly stated as: we should expect

Some, like Alexander Vilenkin (Vilenkin, 2011) and Nick Bostrom (Bostrom, 2002) have argued via principles such as *The Principle of Mediocrity* and *The Self Sampling Assumption* (respectively) that one should assume they are typical and expect to see evidence that corresponds to that of a typical observer. This assumption that we are typical observers is known as the typicality assumption. If we grant the typicality assumption, we can test theories by checking to see if our evidence corresponds to the evidence a typical observer will have. In our toy model, typical observers in the multiverse will be in matter dominated universes according to one theory while according to another theory typical observers will be in antimatter dominated universes. Since we assume that we are typical observers, we can prefer the theory that describes typical observers as being in matter dominated universes since we ourselves are in a matter dominated universe (over the theory that says typical observers are in antimatter dominated universes).

Others, such as Jim Hartle & Mark Srednicki (Hartle and Srednicki, 2007, Hartle and Hertog, 2015), have criticized this position, arguing that we should believe we are typical observers only if we have evidence that indicates we are typical. They argue that since it could be that we are atypical observers in the multiverse, we cannot grant the typicality assumption. Further criticisms like those of Emily Adlam (Adlam, 2024) have maintained that self locating credences (such as credences about how typical we are) are based solely on subjective uncertainty, and as such are not rationally constrained by anything except probabilistic consistency.[5]

In this paper, I wish to achieve two goals:

1. Show that arguments for the typicality assumption put forward by Vilenkin and Bostrom fail to hold in the multiverse scenario.

2. Show the possibility for empirical confirmation in the multiverse without using any assumptions about how typical we are. In doing so, I will disagree with Hartle and Srednicki, who

to observe conditions that are hospitable for life. In our multiverse context, the anthropic principle captures the idea that we should only expect to observe a universe that is hospitable for life, and so in making predictions we should include anthropic considerations that constrain the values of our observables. In this paper, I will assume that all anthropic principle considerations have been taken into account. For example in the matter vs antimatter case, I will assume that life can arise in both matter and antimatter dominated universes, and so we cannot constrain our expectations on this via anthropic considerations. The second topic we will not discuss is the measure problem, which arises mostly in models of inflation that lead to eternal inflation (which most do). Eternal inflation is eternal into the future and spawns an infinite number of universes in the multiverse. Given that the number of universes is infinite, one needs to specify a probability measure to make statements about any kind of property distribution in the multiverse. However, there is considerable debate on what the correct measure should be. See Linde and Noorbala, 2010. In this paper, I assume that a generally agreed upon measure has been specified.

[5] Probabilistic consistency meaning that your distribution of credences does not violate the axioms of probability.

claim that we can extract predictions from multiverse theories only if we assume a measure of how typical we are (called a xerographic distribution). I will also disagree with Adlam who argues that empirical confirmation using self-locating beliefs is just not possible in these cases.[6]

To achieve these goals, my plan for this paper is as follows:

- In §2, I will detail the arguments introduced by Vilenkin and Bostrom for the typicality assumption and criticize the justification given for its adoption by utilizing Adlam's arguments against self-locating beliefs. I will also show that the toy models introduced by Vilenkin and Bostrom to defend the typicality assumption share an important dissimilarity with the multiverse case that renders them unsuitable as an analogy to the multiverse.
- In §3, I argue that there is no need for a typicality assumption in the first place for cosmological theories if we construe our evidence in a 'third person' manner. That is, construing our evidence as 'there is at least one instance of life measuring the value x for some observable' instead of 'we measure the value x for an observable' lets us empirically test theories without assuming anything about our typicality.
- §4 will conclude the discussion and suggest that we can extend the results of this paper to open problems such as the Doomsday Paradox.

2 The Self Sampling Assumption and the Principle of Mediocrity

Notions of typicality have appeared in the literature under the guise of *The Principle of Mediocrity* (Garriga and Vilenkin, 2007) and the *Self Sampling Assumption* (Bostrom, 2002).

Bostrom defines the Self Sampling Assumption (SSA) to be:

> *One should reason as if one were a random sample from the set of all observers in one's reference class* (Bostrom, 2002, pg 57).

The Principle of Mediocrity is defined as follows:

[6] Adlam distinguishes between *pure self-locating* (PSL) credences and *superficially self-locating* (SSL) credences. According to Adlam, PSL credences cannot be used for empirical confirmation while SSL credences may be used. The typicality assumption involves the usage of PSL credences.

> *We would suggest that, on the contrary, we should assume ourselves to be typical in any class that we belong to, unless there is some evidence to the contrary.* (Garriga and Vilenkin, 2007, pg. 3)

Despite the apparent similarity, the Principle of Mediocrity (PM) differs from the Self Sampling Assumption in an important way. For this reason, I will analyze the PM and SSA separately. The overarching goal for this section is to show that neither SSA nor PM holds in the multiverse scenario.

2.1 The Self Sampling Assumption

Bostrom uses a toy model titled DUNGEONS to motivate SSA:

> *The world consists of a dungeon that has one hundred cells. In each cell there is one prisoner. Ninety of the cells are painted blue on the outside and the other ten are painted red. Each prisoner is asked to guess whether he is in a blue or a red cell. (Everybody knows all this.) You find yourself in one of the cells. What color should you think it is?—Answer: Blue, with 90% probability* (Bostrom, 2002, pg 59).

In this model, you are not told what cell you are in, only that 90% of the cells are blue cells. The argument that Bostrom uses to defend the typicality assumption is that it provides the best strategy for the group as a whole. That is, if everyone assumed they are typical and that they are in a blue cell, 90% of the prisoners would win while only 10% would lose. If instead, the prisoners had bet that they are atypical, so that they are in a red cell, then only 10% of the prisoners would win. Thus Bostrom defends typicality by appealing to the best betting strategy for the group. I shall denote this line of reasoning as COLLECTIVE GOOD.

I am not convinced by COLLECTIVE GOOD. It is not clear why we should use the best betting strategy for the group. Why should we care about whether other people get it right or not? After all, we are not other people, we are ourselves. Perhaps what Bostrom means here is that other people getting it right somehow improves our own chances of getting it right. But why other people getting it right would influence our chances of getting it right is not clear. Using the best betting strategy for the group for the group improves your betting strategy only if you assume you are randomly selected from the group. But then you have already assumed what you have set out to motivate.

There is also an issue of specifying the reference class over which one assumes typicality. SSA is defined over a reference class, but how do we decide what this reference class includes? Suppose for example that instead of the dungeon consisting of all human prisoners, 10 blue cells contain polar bears. Should we include the polar bears in our reference class and conclude there is a 90% chance that we are in a blue cell? Or should we exclude polar bears from the reference class and recalculate the probability of finding ourselves in a blue cell as 80/90?

Before we proceed with our discussion, it is important to explicate how Bostrom is understanding the toy model. When we are told that there is a prisoner in each cell, we are meant to understand that we could be any one of these prisoners we are told about. For example, the guard may have told you that there are a hundred people in the dungeon including you and most of them are in blue cells. What you are meant to understand here is that since you don't know which cell you are in, you are uncertain about which person you are. For example, you could be the person in the first cell or the person in the second cell or the person in the third cell, etc. In this manner of thinking, Bostrom can reply to the reference class issue by arguing that one should assert typicality only over observers they could possibly find themselves as. Indeed he does:

> *Now, whether the polar bears count as observers who are members of the reference class makes no difference. Whether they do or not, you know you are not one of them. Thus you know that you are not in one of the ten cells they occupy. You therefore recalculate the probability of being in a blue cell to be 80/90, since 80 out of the 90 observers whom you—for all you know— might be, are in blue cells.* (Bostrom, 2002, pg 70)

For Bostrom, it does not matter who is in the cell as long as you could possibly be that person. If the guard had told us that there is someone named Jerry Smith in a blue cell, and we know we are not Jerry Smith, we aren't to include Jerry Smith in our class used to assert typicality.[7]

Having clarified Bostrom's view, one may wonder whether SSA is reducible to the Principle of Indifference applied to which observer one is. Indeed one would demonstrably get the same results as Bostrom if one were to simply apply the Principle of Indifference over a list of possible observers one could be.

In the dungeons scenario, we do not know which cell we are in, and hence which person we are. For all we know, we could be in any one of the 100 cells. The Principle of Indifference (PoI) suggests that in situations where we have no reason for favoring one possibility over other , we should ascribe

[7]One may wonder what exactly Bostrom means when he says that it does not matter whether polar bears are part of the reference class. Trivially, since SSA is defined over a reference class it clearly matters whether polar bears are in the class. Perhaps what Bostrom means here is that we need not make any arguments appealing to the dissimilarity between polar bears and humans since we know from the outset we are not polar bears.

all possibilities equal probability. [8] Assigning probabilities in accordance with the PoI would give you a 90% chance for finding yourself in a blue cell. If we come to know that some of the prisoners in the blue cells are polar bears, then PoI demands that we do not include those cells in the calculation since we know we are not in those cells. Putatively, PoI seems to give the same results as SSA.

We can also notice that when we talk about 'being any one of the prisoners in the dungeon' what is really meant here is that we could be in any one of the cells. That is, we are not to understand 'we could be any of the prisoners' to mean that it is metaphysically possible for us to be a different person. Rather, it is simply that we could be in any cell in the dungeon. Understanding this point makes it clear that SSA in DUNGEONS is reducible the PoI applied over cells we could be in. Consider the scenario where there are no other prisoners in the dungeon (and we know this). We are told that there are 100 cells in the dungeon (with the same color distribution) and we are in any one of these cells. Now, we do not have any other observers in the dungeon except for us. Intuitively, we may still want to believe that there is a 90% probability for being in a blue cell. However, if SSA is solely based on a reference class of other observers, we cannot get this probability since we are the only observer. The only way to make sense of a 90% probability for blue cell in this case, is to make sense of SSA as an application of the PoI. As further evidence of this, Bostrom himself uses SSA to conduct empirical confirmation in cases where other observers are not clearly specified:

> *consider the two rival theories T1 and T2 about the temperature of the cosmic microwave background radiation. (T1 was the theory that says that the temperature of the cosmic microwave background radiation is about 2.7 K (the observed value); T2 says it is 3.1 K.)* (Bostrom, 2002 Chapter 5, pg 73)

The situation Bostrom is imagining here is one where the CMB does not have the same value everywhere in the universe, rather most of the universe has the mean value but some parts have values that are different from the mean value. One can then ask what reason we have for asserting that the value we measure is the mean value? Bostrom argues that SSA provides the justification for assuming that we have the mean value. He argues for applying SSA across all life in the universe and noticing that most life forms measure the mean value, which justifies our assumption that we have the mean value. However, cosmological theories about the CMB do not describe the existence

[8] The Principle of Indifference (PoI), first coined by Keynes (1921) has been subject to much controversy. There are versions of the PoI such as those of David Builes (Builes, 2024) that apply to cases (like ours) of self-locating uncertainty and putatively seek to resolve some of the most common complaints by requiring the sample space be maximally specific. Still, PoI remains controversial especially when there are an infinite number of possibilities. See Norton (2008)

of other observers. Similar to the dungeons scenario without other prisoners, CMB theories simply describe the value of the CMB at most places in the universe. For SSA to give us the reasoning for preferring T1 over T2 without talking about other observers, we have to understand SSA as just the PoI, and the reference class SSA is asserted over as a reference class consisting of places in the universe we could find ourselves in. Understanding SSA as the PoI would give the desired result by arguing that since we could be anywhere in the universe and most places have a mean CMB value, we ought to assume we also have the mean CMB value.

If SSA is indeed just an application of the PoI in disguise, then arguments against the usage of the Principle of Indifference automatically transfer over to SSA. Fortunately for SSA, since it comes with the appropriate partitions of the sample space predefined (a set of possibilities in one possible world one could be in), it has room to escape the most common criticisms such as Bertrand's Paradox.[9] However, SSA is still susceptible to challenges like those of Adlam (Adlam, 2024) which questions the justification of using PoI for empirical confirmation in all scientific theories. Adlam raises the point that the successful applications of the PoI always correspond to symmetries in the physical process that lead to a partition in the outcome space that is the same as the partition assumed by PoI. That is, success for empirical confirmation using PoI is dependent on the set of outcomes of the physical process possessing symmetries that give probabilities to outcomes that are the same as the probabilities assigned by PoI. For example, there are situations like tossing a fair coin, where the phase space of possibilities has a symmetry that equally partitions heads or tails. But there are also situations where the physical process does not have this property, and the phase space of possibilities is partitioned in an unequal manner (for example an unfair coin toss). As Adlam notes, one must know that the process responsible for the outcomes has the associated symmetries in the phase space in order to use PoI successfully:

> *The key point is that the partition that will yield the most successful results in an application of the NSL-POI is not knowable a priori just from an abstract description of the outcome space—the success of the method depends on how closely the choice of partition reflects the features of the process by which the outcome is selected* (Adlam, 2024, pg. 16)

Like the coin toss, for PoI to be a useful tool for empirical confirmation in DUNGEONS, we need to assume a process that assigns us to a cells randomly. However, we are not told that we have been

[9] Bertrand's paradox is a well known paradox which arises when there are multiple ways to partition the sample space with each way giving a different probability for the same event. See Shackel, 2007, Norton, 2008.

randomly assigned to a cell. It could have been a random process like the guard randomly choosing a cell, or it could have been some deterministic process that always results in us getting assigned to a red cell. Suppose one guard tells you that most of the cells are blue cells and another guard tells you that most of the cells are red cells. Bostrom says that if we find ourselves to be in a red cell, we should believe the second guard over the first one since SSA dictates that the chance for observing a red cell is higher given the second guard's story over the first guard's. This argument only works if you believe that you were assigned to a cell randomly. Suppose that this was not the case and in fact you were assigned to a red cell based on your crime. If there was such a deterministic process that meant that you would always end up in a red cell, then the chances of seeing that you are in a red cell are the same for both guard's stories. So preferring one guard's story over the other using SSA/PoI only works if we assume that we have been randomly put into a cell, which would give the outcome space the necessary symmetry to ensure an equal probability for being in any one cell.

Regardless of whether it is justified to assume that we are randomly assigned to a cell in the dungeons scenario, clearly we cannot claim the same for the multiverse case. There is no sense in which one gets randomly assigned to a universe. Just like nobody assigns us to planets, nobody or no process assigns us to universes. Given that there is no process that randomly distributes us across universes, we cannot use justifiably SSA or PoI as a tool for empirical confirmation. Bostrom's argument that we can assume typicality is thus refuted by noticing that SSA (which justifies the typicality assumption) cannot be used as a tool for empirical confirmation in multiverse cosmology. Assuming typicality in the multiverse would be akin to assuming the second guard's description of the dungeon is the correct one (after learning our cell is red), despite knowing that we were not randomly assigned to a cell.

2.2 The Principle of Mediocrity

The Principle of Mediocrity, as defined by Vilenkin makes a similar claim to SSA, namely that one should assume one is typical and expect to see evidence that a typical observer would see. However, there is a striking difference between Vilenkin's PM and Bostrom's SSA. To spot the difference, remember that the class over which one asserted SSA was a class of observers one could be (read: possible situations one could find oneself in). However in Vilenkin's PM, this is not the case as we can see from his toy model:

> *Imagine that as you arrive to a meeting of the Royal Society, the organizers put a white*

> *or black hat on you. They have removed all mirrors, so you don't know the color of your hat. You notice though that 80% of people around you wear white hats and 20% wear black hats. There may or may not be some system as to how the hats are distributed. For example, the color could be correlated with your sex, age, height, etc - but you don't know.* (Vilenkin, 2011, 5.35)

Notice that in Vilenkin's scenario, the class over which one asserts typicality is defined as all the people at the meeting. SSA will not give this class, since the PoI requires it to be possible that we could be any of the observers in the class. Clearly, it does not make sense to reason that we could be any of the observers in Vilenkin's class since we can simply look around and notice that they are other people and not us. SSA reasoning would require us to remove observers from our class once we know we are not them (like the Jerry Smith example); so in the SSA way of thinking, you remove all the people you see around since it is evident that we could not be those people we see. Thus the class given by SSA consists of just the two situations you could find yourself in: one where you are wearing a white hat and another where you are wearing a black hat. Since we do not know which hat we are wearing, the PoI (and hence SSA) will give a 50% probability for a white hat and the same for a black hat. However, as we will see, Vilenkin thinks that we should believe there is an 80% probability that we have a white hat, clearly different from SSA.

Just like Bostrom, Vilenkin uses COLLECTIVE GOOD as justification for asserting typicality. In Vilenkin's case, the reference class problem makes COLLECTIVE GOOD especially suspect. Imagine if we add some polar bears with hats to the Royal Society. Unlike Bostrom, Vilenkin cannot justify the exclusion of polar bears by arguing that we are not polar bears. Such an argument would also exclude everybody else we see around us and reduce our class to just us.

To proceed with our discussion, it will do us good by way of avoiding red herrings to slightly edit Vilenkin's hats scenario. The proposed edit will change how we come to know that 80% of the members are wearing white hats. Instead of learning this figure by looking around and noticing what color hat other members are wearing, we learn this figure from one of the organizers who has told you that 80% of the members (not including you) have white hats. The advantage of this edit is that it prevents ambiguities to do with how exactly you are ignorant to the rules for sorting. For example, someone might question how exactly it is that we can look around and see other people's hats but maintain that we do not know if the hats are sorted by sex (after all it would seem like in the case the hats were sorted by sex you would instantly be able to tell by looking around). To get around these objections and preserve the genuine uncertainty in the sorting rule, I edit the scenario such that we are just told that 80%(again non-inclusive of you) of the members are wearing white hats and the rest black hats. Importantly, this situation still preserves the distinction from Bostrom's DUNGEONS by the organizers telling you that you were not included in the '80% wear white hats and 20% wear black hats' tally.

Given this edited situation, let us turn to the question at hand. What hat have they put on you? As stated earlier, if we go by the Principle of Indifference, then there are two possibilities corresponding to the two observers you could be: you with a black hat and you with a white hat. Remember that the Principle of Indifference (and hence SSA) can only apply when we have no reason for favoring one epistemic possibility over the other. In this hats scenario, what Vilenkin has to argue is that the empirical evidence of other people's hat color distribution is relevant to your belief about your own hat color. Vilenkin's intuition behind typicality may seem reasonable. In fact I agree with Vilenkin that one can reasonably believe they are wearing a white hat in this situation. However, I disagree with Vilenkin that this scenario is analogous to the multiverse situation such that justification for typicality in one carries over to the other.

The kind of reasoning in which typicality is adopted in the hats scenario involves taking the distribution of other people's hats as evidence that the organizers at the Royal Society have decided to assign hat colors in some particular manner. That is, if you see that 80% of the members are wearing white hats, this is evidence that the organizers have decided to assign hats such that each person has an 80% chance of getting a white hat. Whether this is a fair inference from the information given is an open question. However, for the moment, let us grant Vilenkin this inference. Given this inference, you can believe you have an 80% chance for having a white hat and a 20% chance for having a black hat. Crucially, this line of reasoning is entirely dependent on understanding there to be a process by which you come to have a black hat or a white hat. In this case, this process amounts to the organizers selecting a hat for you (in accordance with the 80/20 distribution). One can easily see that we can assert Vilenkin's typicality only insofar as we believe this process exists. However, clearly there are cases in which such a process does not exist. For example, suppose that there are 80 Billion chimpanzees and 8 billion humans. It seems absurd to conclude that there was a 10% chance that I ended up as a human as opposed to a chimpanzee. Vilenkin's reasoning does not seem absurd because it seems reasonable to say that there is a 20% chance I end up with a black hat and an 80% chance I end up with a white hat. The key element present in Vilenkin's case that makes his case seem reasonable is the organizing committee deciding who gets what hat in accordance with the 80/20 distribution. Unlike the organizers in the Royal Society, no one is assigning people to be a human or a chimpanzee. This is also the case with the multiverse; as we have stated before, there is no process by which you get assigned to be in one universe over others. Since no one is assigning us to universes, assigning probabilities for which universe we are in based on Vilenkin's typicality is akin to assuming there was a 10% chance I was born as a chimpanzee. So clearly, Vilenkin's typicality is not applicable in cosmology.

3 Without a Typicality Assumption

In this section, I wish to show that it is not necessary to make any assumptions about typicality in order to extract predictions from multiverse theories. That is, we need not worry about how typical or atypical we are for evaluating theories of the multiverse. In the first subsection, I will show an alternate way of construing our evidence that lets us perform empirical confirmation without any assumption about typicality. In the second subsection, I will respond to an argument from Roger White, who claims that we cannot construe our evidence in the manner I wish to.

3.1 Is a Xerographic Distribution Necessary?

So far, we have shown that neither SSA nor the PM is suitable in the cosmological setting. But one can reasonably ask how one should evaluate multiverse theories. Different universes have different properties. What we ought to see given a multiverse theory will depend on which universe we are in. If SSA and PM are unjustified in the multiverse setting, what do we do now to conduct empirical confirmation for our theories? To situate our discussion, let us introduce two toy theories:

A: There are 20 bubble universes, which we can label as U1 through U20. U1 and U2 are matter dominated and the rest are antimatter dominated.

B: There are 20 bubble universes, which we can label as U1 through U20. U1 through U16 are matter dominated and the rest are antimatter dominated.

Assume also that we have a theory of life that describes a single human existing in each universe.[10] Vilenkin answers that we can assume that we will see the typical values and seeing non typical

[10] For the purpose of simplicity, we will assume we only have one human in our universe. Nothing in the following discussion relies on the details of our theory of life so I am justified in making this simplification.

values will be evidence against the theory. In our model, seeing a matter dominated universe will be evidence supporting Theory B and undermining Theory A. Bostrom agrees, and states that SSA provides the missing link between observation and prediction from theory in these cases. But as we have seen, this is not a satisfactory answer.

Hartle and Srednicki (Hartle and Srednicki, 2007) adopt an opposing position, arguing that one cannot trivially assume typicality for all scenarios. Following the terminology introduced by Srednicki and Hartle (Srednicki and Hartle, 2010), let us denote the probability distribution for which universe we are in as the 'xerographic distribution' symbolized as ξ. We can think of ξ as a distribution of probabilities over universes where the probabilities represent the chance we could be in that universe. In our model here, the xerographic distribution specifies the probability we are in U1, U2, U3 etc. Vilenkin argues for ξ such that the probability of being in U1 is equal to U2 is equal to U3 etc (probability for being randomly selected). Let us denote this xerographic distribution motivated by the PM as ξ_{M}. Srednicki and Hartle reject the a priori adoption of ξ_{M}, arguing that there are cases where assuming ξ_{M} leads to absurd results that are contrary to standard scientific practice (such as their Humans v Jovians case in Hartle and Srednicki, 2007[11]). Regardless of the disagreements of what ξ should look like, Hartle and Srednicki claim that one must adopt some ξ in order to extract predictions from theories:

> *Without an assumed xerographic distribution, no prediction whatsoever can be made about what 'we' will see in the future.* (Srednicki and Hartle, 2010)

Adlam argues that since PM and SSA fail to be justified, there are no rational constraints one can place on the xerographic distribution (other than probabilistic consistency). And if one cannot provide a rational basis for choosing a particular ξ, then one cannot use it for empirical confirmation:[12]

> *But the arguments given this article suggest a much more general response: it is 'presumptuous' under any circumstances to use PSL credences to arrive at substantive conclusions about*

[11] The humans vs Jovians case is a thought experiment that showcases a situation where the typicality assumption clearly seems wrong. The situation involves two theories, one which says there are is a lot of intelligent life on Jupiter (compared to intelligent life on Earth) and another theory which says there is no intelligent life on Jupiter. If we assume typicality, and we notice that we are on Earth, we are led to prefer the theory that makes Earthlings typical in the solar system, i.e. the theory that says there is no life on Jupiter. Notice that we have preferred the theory that says there is no life on Jupiter based on evidence from Earth, we have not looked at Jupiter at all. This seems contrary to scientific practice.

[12] As stated earlier, Adlam distinguishes between *pure* self locating belief and *superficially* self locating belief. PSL (pure self locating) credences are associated with one possible world, while SSL (superficially self locating) credences are associated with different possible worlds. In the multiverse case, beliefs about which universe you are in are beliefs about one possible world and so classify as PSL credences.

physics or the content of reality, because there is no rationally compelling way to assign PSL credences, so such credences are not a suitable basis for scientific reasoning (Adlam, 2024, pg 34)

In this section, I will disagree with Srednicki, Hartle and Adlam and show that empirical confirmation *is* possible in these scenarios by construing our evidence in a particular way.

Let us take the two theories A and B again. What we want to know is the probability of seeing the evidence given the theory. That is, we want to know $P(E|T_{\mathrm{A}})$ and compare it with $P(E|T_{\mathrm{B}})$ to see which theory we should prefer. What is the evidence? In this case the evidence is that our universe is matter dominated.[13]

E: Our universe is matter dominated.

Clearly, the chances of seeing this evidence will depend on which universe we are in. If we are in U10, the chances of seeing this evidence given Theory A is 0 while Theory B gives us a probability 1 for observing matter dominance. Hartle and Srednicki argue that since the evidence depends on which universe we are in, we have to assume some xerographic distribution to conditionalize the chances of observing our evidence. To show why I disagree, let me state first without defending, one way our evidence can be construed:

E': There is at least one instance of life observing their universe is matter dominated.

Here what I have done is changed the first person evidence E to the third person evidence *E'*. I have removed any self locating reference in our evidence by describing ourselves in a third person way. The difference between *E* and *E'* is not just that one talks about "at least one instance" and the other does not; it is also that one contains a self referential element (by saying *we* make a measurement) and the other does not (it says *life* makes a measurement). If our evidence is *E'*, then we can assess Theory A and Theory B without the need for a xerographic distribution. We simply look at both theories and see what the chances are for at least one instantiation of life observing their universe is matter dominated. Let us see how we can do this for one theory and universe, U1 for example. Let λ_{U1} denote the objective probability that U1 produces observers who have the same evidence as us.

[13] $P(E|T_{\mathrm{A}})$ represents the conditional probability for seeing the evidence E, given T_{A} (Theory A). The same goes for $P(E|T_{\mathrm{B}})$

Here is the way these likelihoods are meant to be understood: λ_{U1} quantifies the probability that U1 would produce observers who arise and see our evidence. Let us assume for a moment that the likelihoods for humans arising are not sensitive to the differences between any of the universes in our list and so $\lambda_{U1} = \lambda_{Ui}$ where i is any one of the matter dominated universes. That is, for all the matter dominated universes, the likelihood of life observing matter dominance is the same and equal to λ_{Ui}, but for all the antimatter dominated universes, the likelihood for life observing matter dominance is trivially 0 (since their universe is antimatter dominated). Now, the chance of at least one instance of life observing a matter dominated universe is simply $1 - (1 - \lambda_{Ui})^n$ where n is the number of matter dominated universes.[14] Theory A predicts that there are 2 matter dominated universes while Theory B says there are 16 of them. Let us plug in some trial values for λ_{Ui} and see where it takes us:

Values of λ_{Ui}	$P(E'\|T_A)$	$P(E'\|T_B)$
0.05	0.098	0.56
0.2	0.36	0.97
0.5	0.75	0.99
0.90	0.99	≈1.00

In this table, I have given the values for the probability of seeing the evidence *E'* given Theory A (in the second column) and Theory B (in the third column). Different rows correspond to different values of λ_{Ui}. As shown, for smaller values of λ_{Ui}, the probability of observing *E'*, is significantly higher in Theory B over Theory A. But for higher values of λ_{Ui}, Theory B and Theory A give almost equal probabilities for observing *E'*. One may wonder whether empirical confirmation is possible in the case where λ_{Ui} is high enough to give near certainty for observing evidence given both theories. But we can rest assured by noticing that λ_{Ui} denotes not just the probability that our evidence is instantiated, but also that life is there to observe that our evidence is instantiated. Since the chance of life arising is thought to be extremely low, we can have reason to believe that most applications will involve low values of the statistical likelihoods. Thus we may conduct empirical confirmation for multiverse theories without making use of a xerographic distribution.

Something to keep in mind is that when we talk about life, what we really mean is observers that could be us. For example, if a theory says our evidence is instantiated in a universe where there are observers but they could not be us (perhaps because they are silicon based or some other differentiating factor), then the statistical likelihoods should give 0 for life seeing that evidence. Our

[14] The term in the brackets denotes the probability that life has not made the relevant observation in that universe (either because life does not arise there or life makes a different observation). Raising it the power of n gives the probability that life has not made the relevant observation in n universes. The probability that at least one instance of life *has* made the relevant observation then becomes 1 minus that term.

definition of life is a third person description of ourselves. If the beings that see the evidence do not match with a third person description of ourselves, then we do not count them as life for the purposes of empirical confirmation.

Construing our evidence as *E'* instead of *E* takes away the important self locating factor that Adlam argues is responsible for the impossibility of empirical confirmation. Since we no longer refer to *our* being in one universe or other, we need not deal with the unconstrained probabilities associated with which universe we are in. We conduct empirical confirmation utilizing only third person evidence, escaping Adlam's worry.

3.2 E vs E': Response to White

As stated earlier, an important part of my argument was the construal of our evidence as *E'* instead of *E*. Such a construal of our evidence is controversial and there are arguments like those of Roger White's, which attempt to show that such a construal falls prey to what's known as the inverse gambler's fallacy.

Suppose that a gambler walks into a casino and sees that a pair of dice are getting rolled. The casino host offers the following two options for the gambler to bet on:

O_1: The dice were previously rolled many times (prior to the gambler walking in).

O_2: The dice were not previously rolled (the roll the gambler saw was the first roll).

After seeing the pair of dice land on double sixes, the gambler reasons that it is super rare to land a double six on your first try but its much more common to land a double six after several tries. Therefore, he bets on O_1 and believes that the dice must have been rolled a bunch of times previously. The gambler is of course mistaken in his judgment since the dice have the same probability for landing on double sixes every roll. The probability for landing on a double six on any particular roll isn't raised by having any number of previous rolls. The fallacy the gambler is committing here is called the inverse gambler's fallacy.[15] Roger White argues that if we construe our evidence as *E'* instead of *E*, then we are committing the inverse gamblers fallacy. To see White's point, assume our gambler has the following two ways of construing his evidence (the dice landing on double sixes):

[15]The regular gambler's fallacy involves the same kind of situation except in the opposite direction. That is, if a gambler after seeing a fair coin land heads five times in a row remarks that the coin is bound to land tails the sixth time, he will be committing the gambler's fallacy. Just like the inverse gambler's fallacy, the gambler's fallacy points out the mistake in assuming the probability for future tails is raised by previous tosses, even though the probability for tails is the same each toss.

E: This particular roll yielded a double six.

E′: At least one roll yielded a double six.

Notice that the probability of seeing *E* is the same given both O_1 and O_2, but the probability of seeing *E'* is higher given O_1 than O_2. That is, since the probability of seeing at least one roll yield a double six increases as the number of rolls increases, the probability of seeing seeing *E'* is higher given O_1. White argues that just like the gambler is mistaken in preferring O_1 over O_2, we are also mistaken in using *E'* instead of *E* to prefer O_1 over O_2:

> *Like the gambler, we have simply witnessed a single big bang producing this universe. And no number of other (universes) can affect the probability of the outcome we observed.* (White, 2000, pg 265)

Now that the we have explained what the inverse gambler's fallacy is, let us take a look at a toy model proposed by White that he claims to map onto the multiverse situation and exposes the fallacy in assuming *E'* over *E*.

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

White argues that if Jane infers that there are several sleepers and dice rollers, then Jane is committing a fallacy. It will pay for us to get clear on the dialectic and explicitly draw out the connections between this case (which I will refer to as PARTNERED JANE) and the multiverse setting to pin down exactly what White is objecting to. Even though White does not state this explicitly, I believe we are to understand that Jane does not know who her partner is, she only knows that she has a partner. This mirrors our epistemic situation in the multiverse where we know we are in a universe but not which particular universe is ours. White originally developed this toy model to use against an argument from fine tuning for favoring multiverse theories. This is why White has Jane only wake up if a really unlikely outcome occurs. Since we are not concerned with fine tuning, we need not focus on that part. Notice that Jane only wakes up if her partner rolls a double six, so it does not matter whether other partners roll a double six. This is why White can object to Jane assuming that there were many dice rollers, her chance of waking up is solely dependent on her particular partner rolling a double six. What other partners roll or don't roll will not affect her chances for observing a double six. So White argues that assuming many dice rollers does not raise the probability of seeing the evidence *E*. According to White, we should use *E* instead of *E'* because it is not sufficient for 'some partner' to roll a double six to ensure Jane wakes up. That is, White (correctly) argues that our evidence follows from *E* but not *E'*.

White thinks our multiverse setting is analogous to PARTNERED JANE. Like Jane, it is not sufficient for some universe to have life that sees our data to ensure that we ourselves see the data:

> *It is certainly not sufficient for us to exist in some universe β, that β is fine-tuned, or even that β is qualitatively exactly as a actually is. After all, if we postulate enough universes, the chances are that there exist several life-permitting uni verses, perhaps even universes with precisely the same initial conditions and fundamental constants as our universe, and containing human beings indistinguishable from us. But we do not inhabit these universes, other folks do.* (White, 2000, pg 268)

White argues that because the multiverse case is in this way analogous to PARTNERED JANE, we should reject *E'* instead of *E* for the same reasons as those in PARTNERED JANE. However, I disagree that PARTNERED JANE is analogous to the multiverse setting. In PARTNERED JANE, there are qualitative differences between the epistemic states instantiated by different people who are woken up from different partners. For example, suppose Bob is also in the experiment. Jane knows that she is not Bob. Before Jane is put to sleep, if Jane is told that some partner will roll a double six, she cannot infer that she will be woken up because 'some partner' could be Bob's partner. This is why Jane can say that *E'* does not ensure that she herself wakes up. However, in the multiverse, we cannot differentiate ourselves from other life in the same way Jane can differentiate herself from Bob. Consider an advanced and extremely detailed cosmological theory that talks about humans in multiple universes with mental states that are subjectively indistinguishable from ours. In this setting, there is no way for us to distinguish between ourselves and these other instances of life. Whatever we can mean via a third person description of ourselves is also instantiated in these other universes. Unlike Jane who can say that she is not Bob, we cannot say we are not these other humans because there are no qualitative differences between us and them. If we cannot say that we are not these other instances of life, then we have to take seriously the possibility that we could be them. Jane knows that Bob's partner rolling a double six will not wake her up, but we do not know that 'some life in some universe' isn't us in our universe. This dissimilarity between the multiverse setting and White's toy model makes the toy model inappropriate as an analogy to the multiverse. And so, White's arguments against *E'* does not hold in the multiverse context.[16]

4 Conclusion

In this paper I have shown two things:

1. SSA and PM are not justified in the multiverse setting so we cannot use the typicality assumption that arises from their application for empirical confirmation.

2. Nevertheless, there is a way to construe first person evidence as third person evidence to conduct empirical theory confirmation. Once we construe our evidence as a single instance of a third person description of ourselves, we can conduct empirical confirmation.

In the future, I would like to extend this discussion to the 'Doomsday Paradox' (Olum, 2002) where an assumption of typicality (like the Self Sampling Assumption or the Principle of Mediocrity) is made use of to generate the paradox. The doomsday paradox in a condensed form, suggests that since the human population is increasing exponentially into the future, most humans will be born closer to the end of humanity rather than in the middle or beginning. Since most humans are born near the end of humanity, SSA or PM might require us to believe that we are typical humans born near the end of humanity and hence doomsday is imminent. However, as the results from this paper conclude, this will only be valid in so far as one can assume a process that randomly distributes your birth rank. Whether we can understand such a process to exist can be the subject of a future paper.

[16]To make the multiverse setting more analogous to PARTNERED JANE, perhaps we can amend the toy model such that all the sleepers are duplicates of Jane with subjectively indistinguishable mental states. In this setting, before Jane goes to sleep, if she is told that 'someone rolled a double six', then she can infer that someone who thinks they are Jane, will be woken up. Jane cannot differentiate between herself and the duplicates the same way she can between herself and Bob. In this scenario, it can perhaps be shown that Jane's evidence ought to be *E'* and that this scenario is analagous to the multiverse setting. If this can be shown, then we have further support for choosing *E'* over *E* in the multiverse context.

References

[1] Emily Adlam. "Against Self-Location". In: *Philosophy of Science* 82.5 (2024), pp. 1069–1081. DOI: 10.1086/732908.

[2] Andreas Albrecht and Paul J. Steinhardt. "Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking". In: *Physical Review Letters* 48 (1982), pp. 1220–1223. DOI: 10.1103/PhysRevLett.48.1220.

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

[4] David Builes. "Center Indifference and Skepticism". In: *Noûs* 58.3 (2024), pp. 778–798. DOI: 10.1111/nous.12489.

[5] Brandon Carter. "Large Number Coincidences and the Anthropic Principle in Cosmology". In: *The Role of Cosmology in the Development of Modern Physics* 6 (1974), pp. 139–139.

[6] Jaume Garriga and Alexander Vilenkin. "Prediction and explanation in the multiverse". In: *arXiv preprint arXiv:0711.2559* (2007).

[7] Alan H. Guth. "The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems". In: *Physical Review D* 23.2 (1981), pp. 347–356. DOI: 10.1103/PhysRevD.23.347.

[8] James Hartle and Thomas Hertog. "The Observer Strikes Back". In: *arXiv preprint arXiv:1503.07205* (2015). DOI: 10.48550/arXiv.1503.07205.

[9] James Hartle and Mark Srednicki. "Are We Typical?" In: *Physical Review D* 75.12 (2007), p. 123523. DOI: 10.1103/PhysRevD.75.123523.

[10] John Maynard Keynes. *A Treatise on Probability*. 1921.

[11] Andrei Linde. "A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy, and Primordial Monopole Problems". In: *Physics Letters B* 108.6 (1982), pp. 389–393. DOI: 10.1016/0370-2693(82)91219-9.

[12] Andrei Linde. "A Brief History of the Multiverse". In: *Reports on Progress in Physics* 80.2 (2017), p. 022001. DOI: 10.1088/1361-6633/aa50e4. arXiv: 1512.01203 [hep-th]. URL: https://arxiv.org/abs/1512.01203.

[13] Andrei Linde and Mahdiyar Noorbala. "Measure problem for eternal and non-eternal inflation". In: *Journal of Cosmology and Astroparticle Physics* 2010.09 (2010). DOI: 10.1088/1475-7516/2010/09/008.

[14] John D. Norton. "Ignorance and Indifference". In: *Philosophy of Science* 75.1 (2008), pp. 45–68. DOI: 10.1086/587822.

[15] Ken D. Olum. "The Doomsday Argument and the Number of Possible Observers". In: *The Philosophical Quarterly* 52.207 (2002), pp. 164–184. DOI: https://doi.org/10.1111/1467-9213.00260.

[16] Nicholas Shackel. "Bertrand's Paradox and the Principle of Indifference". In: *Philosophy of Science* 74.2 (2007), pp. 150–175. DOI: 10.1086/519028.

[17] Mark Srednicki and James Hartle. "Science in a Very Large Universe". In: *Physical Review D* 81 (2010), p. 123524. DOI: 10.1103/PhysRevD.81.123524.

[18] Leonard Susskind. *The Anthropic Landscape of String Theory*. 2003.

[19] Alexander Vilenkin. *The Principle of Mediocrity*. 2011.

[20] Roger White. "Evidential Symmetry and Mushy Credence". In: *Oxford Studies in Epistemology* 3 (2000), pp. 161–186.