Anthropic reasoning in multiverse cosmology and string theory*

Steven Weinstein†

Perimeter Institute for Theoretical Physics, 31 Caroline St, Waterloo, ON N2L 2Y5 Canada
Dept. of Philosophy, University of Waterloo, Waterloo, ON N2L 3G1 Canada
Dept. of Physics, University of Waterloo, Waterloo, ON N2L 3G1 Canada

July 31, 2005

Abstract

Anthropic arguments in multiverse cosmology and string theory rely on the weak anthropic principle (WAP). We show that the principle, though ultimately a tautology, is nevertheless ambiguous. It can be reformulated in one of two unambiguous ways, which we refer to as WAP_1 and WAP_2. We show that WAP_2, the version most commonly used in anthropic reasoning, makes no physical predictions unless supplemented by a further assumption of "typicality", and we argue that this assumption is both misguided and unjustified. WAP_1, however, requires no such supplementation; it directly implies that any theory that assigns a non-zero probability to our universe predicts that *we* will observe our universe with probability one. We argue, therefore, that WAP_1 is preferable, and note that it has the benefit of avoiding the inductive overreach characteristic of much anthropic reasoning.

1 Introduction

Over the last twenty years or so, inflationary cosmologists have been toying with theoretical models that postulate the existence of a multiverse, a set of quasi-universes (henceforth "universes") which are more-or-less causally disjoint [1],[2]. String theorists, who once hoped to predict a unique supersymmetric extension to the standard model, now find themselves contemplating a similar scenario, with perhaps 10^{500} or more metastable low-energy vacua (the "landscape") realized via inflationary mechanisms as essentially distinct universes [3],[4],[5],[6]. The vast majority of the universes in these scenarios look nothing like our universe, the values of the fundamental physical parameters (e.g., the cosmological constant) differing markedly from the values we observe.

A significant number of physicists have understood this situation — in particular, the fact that universes like ours are atypical — to be problematic because, it is felt, it reveals a lack of explanatory or predictive power. Others seem more concerned with our inability to either confirm or falsify such theories. It is in addressing these concerns that anthropic reasoning enters the picture.

*Thanks to Rob Caldwell, Jim Peebles, and Lee Smolin for helpful discussions and comments on an earlier draft.

†Email: sw@uwaterloo.ca

Anthropic *reasoning* is reasoning based on the use of the (weak) anthropic *principle*, articulated by Carter as

> **WAP**: "What we can expect to observe must be restricted by the conditions necessary for our presence as observers." ([7], p. 291).

This seems unobjectionable; in fact, it is a tautology. As such, one wonders how it can do any explanatory or methodological work, much less be the object of heated disagreement. We will argue that the anthropic principle, as stated by Carter, contains what Bohr might have called an "essential ambiguity", and that the problematic aspects of anthropic reasoning stem from the use of a form of the principle which requires recourse to an *additional* and unsupported assumption of "typicality" in order to make contact with observation. Though the crucial importance of the "principle of mediocrity" which encodes the assumption of typicality is understood by some, its connection to a particular form of the WAP and its status as a distinct and unargued-for assumption seems not to be well understood.

2 Anthropic arguments

2.1 General

The general strategy of an anthropic argument is as follows. Begin with some mulitiverse hypothesis, a hypothesis which gives:

- a set of possible universes parametrized by the values taken by the cosmological constant, the dark matter density, and other physical parameters, and
- a probability distribution describing the relative frequency of occurrence for these universes in the (generally infinite) ensemble.

Now, in accordance with WAP, restrict attention to the subset of universes that support the existence of observers, since these are the only universes we could hope to observe, and consider the distribution of the parameters over this "anthropic" subset. These are the universes it is possible to observe, and so the predictions for what we can expect to observe should be based on this subset.

Given the restriction to the anthropic subset, one has a new probability distribution for each parameter reflecting the relative frequency with which the various values of the parameter occur in the anthropic subset. If the parameters take on a continuous set of values, or even a large but finite number, the probability of any particular value will be absurdly small, so that no particular outcome is to be expected. What, then, does anthropic reasoning predict?

It is here that one must introduce an assumption of typicality, such as Vilenkin's "principle of mediocrity," which he describes as "the assumption that we are typical among the observers in the universe." He elaborates,

> Quantitatively, this can be expressed as the expectation that we should find ourselves, say, within the 95% range of the distribution. This can be regarded as a prediction at a 95% confidence level. ([8], p. 2)

Equipped with this further assumption, one can infer that the observed values of the parameters will lie within the 95% range.[1] In short, if we assume we are typical (i.e., if we apply the principle of mediocrity), we get a range of predicted outcomes which is a proper subset of the anthropic subset.

From here, the anthropic reasoner will generally conclude that the observation of an (anthropically) typical value by us means that the theory has successfully *explained* the value (since we are a generic *prediction* of the theory). Furthermore, she will judge that the observation of a typical value offers inductive *support* to the theory.

Similarly, observation of an *a*typical value is judged to constitute a disconfirmation, even falsification, of the theory [9],[10]. However, from a logical standpoint, all that has been shown by a failure to observe an anthropically typical value is that the *conjunction* of the theory and the principle of mediocrity is inadequate. One might just as well impugn the principle of mediocrity rather than the theory. The following example brings this out.

2.2 Example: The Googolverse

Consider a multiverse theory which postulates the existence of a "Googolverse", an infinite ensemble made up of 10^{100} (a googol) different sorts of universe, where "different sorts" means that the universes are characterized by different values of one or more physical parameters. The postulated theory also provides a probability distribution over the parameters, so that one can talk about "typical" and "atypical" values. Suppose, finally, that the values *we* observe are atypical, lying outside the 95% confidence interval. In such a situation, we might turn to the anthropic principle in order to assess the theory.

Upon doing an anthropic analysis of this multiverse, let us suppose we discover that 10^5 of the 10^{100} universes support observers. In accordance with the anthropic principle, we reason that these are the only universes we could possibly observe, and we go on to ask whether what we observe — the value of the cosmological constant, the electron mass, and so on — is typical of this observer-supporting, "anthropic" subset. In other words, would a generic observer expect to see something like what we see? The anthropic reasoner says that if the answer is no, then our theory should be regarded as an explanatory failure, and our observation of the atypical values should count against acceptance of the theory.

So far, so good... perhaps. But suppose we analyze the observer-supporting subset further, and discover that there are really only two general sorts of observers, Humans and Aliens (the reader is encouraged to envision insect-like creatures akin to those portrayed in the film *Alien*), and that Aliens are by far the dominant form of life in the observer-supporting subset. In fact, we discover that only 17 of the 10^5 sorts of universes support the existence of Humans (Aliens are far more adaptable, apparently), and that these account for only 0.1% (i.e., 100) of the universes in the observer-supporting subset. Statistical analysis reveals that of this Human-accommodating subset of 17 universes, the values we observe in our universe, one of the 17, are entirely typical.

In short, we come out looking like an atypical member of the Googolverse, an atypical member of the observer-supporting subset, and a typical member of the Human-supporting subset. Most anthropic reasoning seeks typicality (or "mediocrity") amongst the broadest

[1] But note that the notion of typicality, of lying within some confidence interval, is not well-defined for all probability distributions. For example, a flat distribution has no "typical" values. What, for example, is a typical outcome of the roll of a fair die?

possible subset of observers, judging that "it is prudent to condition probabilities, not on a detailed description of 'us', but on the weakest condition consistent with 'us' that plausibly provides useful results." ([11], p. 5). Thus most anthropic reasoners would presumably conclude that the observed values are not *explained* by the Googolverse hypothesis, and *a fortiori* that these values do not *support* the hypothesis. On the other hand, the typical member of the small subset of the human race who is not completely flummoxed by the idea of a multiverse might well disagree, noting that the Googolverse hypothesis succeeds very well in explaining what *we* observe, since *we* are Humans, not Aliens.

3 The Essential Ambiguity

The Googolverse example is of course a caricature, but it highlights the way in which the particular choice of a reference class can make an enormous difference to the outcome of an anthropic argument. It does so via the additional requirement of typicality with respect to the reference class of observers.[2] Given the seemingly uncontroversial nature of the WAP, it is surprising that its application can be so contentious. In this section, we will see that this is a result of an ambiguity in the principle itself.

Consider once again the Weak Anthropic Principle, "What we can expect to observe must be restricted by the conditions necessary for our presence as observers." How are we to understand the phrase "our presence as observers"? Are we to understand it as talking about *our* presence, or the presence of observers, *in general*. The difference is pivotal.

Given that we *are* observers, the conditions necessary for our presence as observers are no different from the conditions necessary for our presence, and we might reformulate the WAP accordingly:

> **WAP**$_1$: "What we can expect to observe must be restricted by the conditions necessary for our presence."

This is an uncontroversial claim, since "we" can only observe the properties of worlds that allow our presence. So for example, we cannot observe a world or a universe in which we failed to evolve, even if that universe has earth-like planets and other DNA-based life forms.[3]

Clearly, most proponents of anthropic reasoning understand Carter's principle in a different way, interpreting WAP as,

> **WAP**$_2$: "What we can expect to observe must be restricted by the conditions necessary for the presence of observers."

Application of WAP$_2$ requires one to establish what "observers" are, and then to identify the conditions necessary for their presence. The constraint is much looser than the constraint imposed by WAP$_1$, in that it admits universes which do not contain "us", do not contain human beings at all. Furthermore, although WAP$_2$ tells us that what we can expect to observe is *restricted* by the conditions necessary for observers, it does not tell us anything about the likelihood of *our* observing particular conditions, even though it does instruct us to calculate the probability distribution of the parameters over the sub-ensemble of universes that support

[2] See [12] and [13] for further discussion of the problem of the reference class.

[3] Interpreting the weak anthropic principle in this way more or less implies the "top-down" approach of [14].

the existence of observers. In order to extract concrete predictions from WAP_2, it must be supplemented by the principle of mediocrity (or something similar), which stipulates that we are *typical* observers. But of course we need not be, and furthermore we may be typical of one class of observers and atypical of another (as in the Googolverse example above).

Despite the vagueness of WAP_2, one might prefer it to WAP_1, since it suggests that we condition on a broader (but unspecified) reference class. However, WAP_2 must be supplemented by the principle of mediocrity, a principle which, unlike other typicality assumptions such as the Copernican principle and the cosmological principle, is invoked not as a working hypothesis but as a principle of reasoning *about* one's working hypothesis (the multiverse theory).[4] Its failings in this regard are readily seen from its application to the Googolverse scenario above. In short, WAP_2 is a kind of fuzzy generalization of WAP_1 which nevertheless leads to no new predictions unless supplemented by the principle of mediocrity, a principle which lacks any independent justification in the abstract and is often wrong in the concrete.

4 Conclusion

WAP_1 and WAP_2 are both interpretations of the weak anthropic principle, and both are incontrovertibly true. Why, then, have we suggested that one should use WAP_1 rather than WAP_2? Because the WAP_1 directly yields testable predictions, while WAP_2 requires one to both identify a particular class of observers and apply the principle of mediocrity with respect to the class chosen. To be sure, WAP_1 makes the simultaneously trivial and strong claim that if the theory assigns a non-vanishing probability to the parameter values we do observe, then that is what we should *expect* to observe. If on the other hand the theory assigns a probability of zero to the values that we observe, then of course the theory is ruled out.

With respect to theory confirmation, one should certainly not say that the fact that the universe we observe is predicted with some probability (however small or large) serves as a confirmation of the theory. But then, why would anyone expect that one could reason inductively from a single data point in the first place?

References

[1] A. Vilenkin. Birth of inflationary universes. *Phys. Rev.*, D27:2848–2855, 1983.

[2] A. Linde. Eternally existing self-reproducing chaotic inflationary universe. *Phys. Lett. B*, 1986.

[3] R. Bousso and J. Polchinski. Quantization of four-form fluxes and dynamical neutralization of the cosmological constant. *J. High Energy Phys.*, 06:006, 2000. hep-th/0004134.

[4] L. Susskind. The anthropic landscape of string theory. 2003. hep-th/0302219.

[5] S. Kachru, R. Kallosh, A. Linde, J. Maldacena, L. McAllister, and S. Trivedi. Towards inflation in string theory. *J. Cosmol. Astropart. Phys.*, 2003(10):013, 2003. hep-th/0308055.

[4] See [15] for an excellent discussion of the Copernicus, Kant, and the anthropic principle, and see [16] for discussion of the roles that the Copernican and cosmological principles have historically played in cosmology.

[6] B. Freivogel, M. Kleban, M. Rodriguez Martinez, and L. Susskind. Observational consequences of a landscape. 2005. hep-th/0505232.

[7] B. Carter. Large number coincidences and the anthropic principle in cosmology. In M. Longair, editor, *Confrontation of cosmological theories with observational data; Proceedings of the Symposium, Krakow, Poland, September 10-12, 1973*, pages 291–298, Dordrecht, 1974. D.Reidel.

[8] A. Vilenkin. Anthropic predictions: The case of the cosmological constant. 2004. astro-ph/0407586.

[9] M. Rees. Numerical coincidences and 'tuning' in cosmology. In N.C. Wickramasinghe, G.R. Burbidge, and J.V. Narlikar, editors, *Fred Hoyle's Universe*, pages 95–108. Kluwer, 2003. astro-ph/0401424.

[10] P.C.W. Davies. Multiverse cosmological models. *Mod. Phys. Lett.*, A19:727–744, 2004. astro-ph/0403047.

[11] J.B. Hartle. Anthropic reasoning and quantum cosmology. 2005. gr-qc/0406104.

[12] N. Bostrom. *Anthropic Bias: Observation Selection Effects in Science and Philosophy.* Routledge, New York, London, 2002.

[13] L. Smolin. Scientific alternatives to the anthropic principle. 2004. hep-th/0407213.

[14] A. Aguirre and M. Tegmark. Multiple universes, cosmic coincidences, and other dark matters. *J. Cosmol. Astropart. Phys.*, 2005(01):003, 2005. hep-th/0409072.

[15] S. Roush. Copernicus, Kant, and the anthropic cosmological principles. *Stud. Hist. Phil. Mod. Phys.*, 34:5–35, 2003.

[16] G.F.R. Ellis. Cosmology and verifiability. *Q.Jl R. astr. Soc.*, 16:245.