Aporia vol. 22 no. 2–2012

Multiverse Theory and the Problem of Induction

Todd DeRose and Daniel Kokotajlo

This paper explores the problem of induction as it applies to multiverse theories. Above and beyond traditional problems of induction—which demonstrate only that our inductive beliefs lack justification—the problem of induction for multiverse theories appears to show that, under such theories, beliefs formed by induction are probably false. The central idea is that a multiverse theorist is obligated to assign equal epistemic weight to every possible world that he could be in, given his experiences so far. Unfortunately, many (if not the vast majority) of these worlds will soon violate his inductive expectations, suggesting that the multiverse theorist should expect the world to dissolve into chaos at any moment.

In section I we go over terminology and outline some key concepts. In section II we give an overview of the traditional problem of induction and some of its more modern variants. In section III we describe in more detail what is common to multiverse theories of the sort we are considering, giving a few examples and a brief account of what motivates philosophers to subscribe to them. In section IV we present our main argument for the conclusion that induction is irrational under multiverse theories. In section V we address several possible objections. section VI is devoted to the objection we think is most crucial, having to do with infinities and

Todd DeRose is majoring in philosophy and theology at the University of Notre Dame. His interests include philosophy of religion, philosophical theology, and metaphysics generally. After graduation, he plans to go on to graduate school in philosophy.

Daniel Kokotajlo is majoring in philosophy at the University of Notre Dame. His interests are broad, but particularly feature theories of everything, most notably multiverse theories. After graduation he plans to get a doctorate in philosophy.

Occam's Razor. Here we will also explain why our argument applies to multiverse theories in ways that it does not apply to non-multiverse theories.

Section I: Key Concepts and Terminology

We think it is important to briefly go over some key concepts and terms before getting started:

> **Multiverse Theory:** Any theory which holds that everything we normally think of as merely logically possible is in fact real, though not necessarily observable by us.
>
> **Phenomenological Counterpart**: Distinct from Lewisian and other types of counterparts, but close to Elga's notion of a similar centered world, a Phenomenological Counterpart to an observer *X* is a being in another world or region of a world which has identical (or at least indistinguishable) phenomenological experiences to *X* up until a given time.
>
> **Chaotic World/Observer**: Any world which, from the perspective of a given observer, remains consistent with that observer's inductive expectations up until a given time, at which point the observer experiences obvious violations of these inductive expectations.
>
> **Sane World/Observer**: Any world which, from the perspective of a given observer, remains consistent with that observer's inductive expectations up until a given time, at which point the observer continues to have experiences in line with these inductive expectations.

Section II: Previous Formulations of the Problem of Induction

It is worth briefly describing some of the problems of induction which have been formulated previously. Obviously there is a vast body of material which has been written on these problems, but we do not think that our argument hinges upon the success or failure of these versions.

The original Humean formulation of the problem of induction (though not by that name) runs roughly as follows:

We believe that the future will resemble the past, in that causal relationships observed previously will continue to apply to similar circumstances in the future—we expect dropping an object will result in it falling, because in the past it has always fallen when we have dropped it. Yet we must ask in what manner this belief is justified. It is always logically possible for a pattern of observations to be broken, no matter how strong or consistent the pattern has been, and so philosophy has had difficulty producing any deductive reason to justify induction. The only apparent alternative is to justify our inductive beliefs through induction, but this is circular. In the absence of a third option for justification (or arguments against the above reasoning), the problem seems to be conclusive: induction has not been justified (Hume section 7; Lewis 159).

More recently, Nelson Goodman has posed the so-called "New Riddle of Induction", a similar problem having to do with our expectations about the future, which goes like this:

It so happens that every emerald we have so far dug up has been green. But every emerald which we have observed could just as well be 'grue'. 'Grue' is defined as "Green during or before (for example) 2014, blue otherwise." We must therefore ask why we assume that emeralds are green rather than grue. It seems we have just as much evidence for the grue hypothesis as we do for the green hypothesis, and attempts to show that properties like 'green' are intrinsically more proper than properties like 'grue' tend to be very ad hoc (Quine 114). Green and grue are just examples; similar reasoning can be applied to all other properties, and so Goodman's problem is quite extensive.

Section III: Examples of Multiverse Theories and their Motivations

Nozick has argued that multiverse theories, as understood in our definition above, have significant advantages over other theories when it comes to ontology (115). Consider: either everything possible exists, or some things exist and other things don't. If the latter, then an explanation is needed for why some possible things exist while others don't. Multiverse theories have the advantage that they do not seem, at least ostensibly, to have to contend with this question—such theories do away entirely with merely possible entities.

Attempting to make better sense of ontology is not the only reason why people have adopted multiverse theories of various kinds. David Lewis, for example, argues in favor of modal realism, a "philosophical multiverse theory" which states that there exists a spatio-temporally and causally isolated world for each logically possible configuration, (excluding those ruled out by the law of noncontradiction). Modal realism, he claims, also makes better sense of our use of ordinary language and modal operators than competing theories do (Plurality).

There are also 'mathematical' multiverse theories which attempt to understand reality through pure abstraction. Max Tegmark's Ultimate Ensemble Theory holds that everything that exists is assembled out of 'mathematical structures', roughly defined as "abstract objects in relation [with each other]" (7). The idea is that if abstract objects (such as numbers and sets) and every possible combination of abstract objects exist necessarily in some sense, and everything we think of as existing in the normal sense can be built out of these abstracts, then it follows that these abstracts also exist in reality. Tegmark's UET promises to reduce literally everything to mathematics, which some may see as an additional help to ontology. For our purposes, Tegmark's theory is not significantly different from Lewis' modal realism or any other forms of multiverse theory as we have defined the term.

Finally, a working multiverse theory might be a good explanation for the apparent 'fine-tuning' of our universe. If every possible universe exists, there must be universes that are tuned for life. Invoking anthropic reasoning, it is only natural that we should find ourselves in one of these universes: If every possible universe exists, then our existence is necessary, and in order to exist we must be in a universe suitable for life.

Section IV: The General Argument

Argument One: Why We Shouldn't Weight Some Worlds More than Others

It follows from the multiverse theories we are considering that for any given observer at any given time, there is a infinite array of worlds which contain observers whose experiences are indistinguishable from those of said observer at least up until that time.[1] Naturally, the same is true of us: we each have an infinite number of counterparts in other worlds (or distant reaches of our own world) who have so far lived identical lives.

This of course does not mean that the worlds these counterparts inhabit are all identical in other ways; presumably, plenty of things could be different about the world without affecting the experiences that an observer has. There might in one world be a star in the far reaches of space that is beyond the ability of the observer to experience, or alternatively (following an argument of Peter Forrest), these universes might contain

[1]It doesn't work because this type of causation does not restrict at all the range of things that could happen. As Lewis says, "Graft any future onto any past" (Plurality, 116).

varying numbers of so-called 'epiphenomenalons' which do not impact our experiences in any way (458). The question, then, is how we are supposed to decide what world we are actually in—which of the many observers in the multiverse is us?

We think that it is fairly intuitively obvious that we have no reason to believe we are in any one universe as opposed to any of the other worlds indistinguishable so far from our own. Obviously, empirical evidence isn't going to give us this information, since the empirical evidence for each possibility is the same. Neither does it seem that a priori reasoning could provide an answer, even in principle. In a multiverse framework every world exists necessarily; thus, it is unclear how deduction could possibly tell us which world or worlds to favor epistemically, as each world is on an equal metaphysical footing. This idea will be elaborated in Section VI.

So far we have discussed negative reasons why we should give equal epistemic weight to each of the worlds that we could be in; we have argued that it would be unreasonable to do otherwise. Now, we will give a positive argument for this conclusion. It is an intuition pump of sorts, and it works by taking a base case in which our claim should be obvious, and subjecting it to a series of intuition-preserving transformations until we have the general result that we want:

Imagine first that you are put to sleep on Sunday, that you are awakened on Monday, that your memory is then erased as you are put back to sleep, and that you are re-awakened on Tuesday under the exact same conditions. When you are awakened and asked what day it is, it seems that you ought to give Monday and Tuesday equal epistemic weight.[2]

Now imagine instead that a Parfit-style teleportation device destroys your present body and reconstitutes two exact copies of you in two distinct (but externally identical) locations.[3] Again, it seems that you should give equal epistemic weight to the two possible locations that you could be at, regardless of the time and distance involved.

Now modify this scenario so that one of your clones is reconstituted in a region of space which obeys the traditional laws of physics, and the other is reconstituted in a region of space where the laws are different (or more complex) but the difference is imperceptible for the time being. The same reasoning applies.

Now remove the teleportation device from our thought experiment, and simply imagine that two identical copies of you have materialized in

[2] See also Elga (Self Locationg Belief) and Lewis (Sleeping Beauty). Both authors concur on this point.

[3] The teletransporter discussed by Parfit is a fairly familiar concept in the literature. See Parfit, Reasons and Persons. Oxford University Press, 12 Apr 1984

these different regions of the universe and you do not know which one you are. Again, you ought to weight each possibility equally.

Now imagine that instead of distant regions of a single universe, we are dealing with spatio-temporally isolated worlds where the possible 'places' you could be are not even connected by space and time. Epistemically, this does not seem different from the prior examples in any meaningful way. However, we have now arrived at the exact situation that multiverse theorists find themselves in every day. So long as each of the worlds in question (that is, chaotic and sane worlds) are phenomenologically indistinguishable from one another up until the present moment, we ought to treat them as equally likely, or at least place the same degree of confidence on each of them.

Argument Two: Why Chaotic Worlds/Observers are at Least as Numerous as Sane Ones

The basic intuition here is fairly straightforward. Imagine that you are dropping your cell phone on the floor. Upon releasing the phone from your grasp, there are a variety of logically possible things that could happen. It could fall down in various paths through the air—or it could shoot sideways, fall upward, hover in place, disappear entirely, or transform into bacon. Under multiverse theories like Lewis' and Tegmark's, there are worlds which are indistinguishable up until the moment at which you release the phone—worlds that, for all you know, you inhabit—in which precisely these things happen. Surely the range of things which could happen that you do not expect given your past experiences is far greater than the range of things that could happen that you would expect. In other words, we might say that the range of possible futures consistent with induction that we could shortly be experiencing is but a narrow band in the full range of logical possibility.

At the very least, it seems that for every sane world we can imagine, there is at least one chaotic world corresponding to it. For example, for every possible way that the cell phone could fall downward when released, there is an 'inverse' way in which it could fall upward, in an opposite path. This alone should lead us to the conclusion that there are at least as many chaotic worlds as sane worlds in the multiverse. Given a one-to-one correspondence between sane worlds and this one particular class of chaotic worlds, we know that the cardinality of the set of all chaotic worlds is the same or greater than the cardinality of the set of all sane worlds.

Now consider Forrest's Principle of the Universability of Rationality. Forrest states that "what is rational for one person is rational for another person who is the same in all relevant mental respects and, who in particular, has the same sources of information" (460). From this it follows that you should choose your beliefs as though you were choosing beliefs for all of your phenomenological counterparts—what is rational for you to believe is also rational for them to believe, and vice versa. This seems to us to be an uncontroversial principle. But surely the belief "I am in a sane world" isn't rational, for it is false for at least as many people as it is true. Thus, it is irrational for us to base our beliefs on induction; we ought rather to expect the world to descend into a chaotic soup of literally unpredictable occurrences, for that is what most of our phenomenological counterparts are going to experience.

In summary, if we accept arguments one and two, then the conclusion that we ought to believe that we live in a chaotic world follows readily. Given that there are at least as many chaotic worlds as sane worlds, and that we should weight each individual world the same, we ought to give as much epistemic weight to the set of all chaotic worlds than to the set of all sane worlds. More concretely, next time we drop our cell phone, we should not be any more surprised if it flies up into the sky than we would be if it smashes into the ground.

Section V: Minor Objections and Responses

There are three objections that we deem minor because they can be easily and decisively answered. Originally we explored them in depth, but because of space constraints we have cut much of this discussion from the paper. Here, we will content ourselves with an extremely brief overview of the three minor objections and our responses:

Anthropic Reasoning Objection: Most chaotic worlds result in the observer dying so fast that he or she doesn't even have time to notice, whereas most sane worlds (in most circumstances) result in the observer living. Thus if we invoke anthropic reasoning (or perhaps more specifically, a psychological-continuity view of consciousness of some sort) and rule out as possibilities those worlds in which we immediately cease to exist, perhaps we will find that the chaos worlds are less numerous than the sane after all.

Response: Even if we allow this ruling-out to take place, we can still find just as many chaotic worlds as sane worlds.

Laws of Nature Objection: Perhaps the regularities we observe aren't mere coincidences—perhaps they stem in some fashion from underlying causes, Laws of Nature for example. Indeed perhaps this is the best explanation for those regularities. Doesn't this give us reason to believe that the regularities will continue to be observed?

Response: Consider the possibility of Laws of Nature that are akin to the property 'Grue:' they dictate that one thing happen in circumstances that we have experienced so far, but dictate that a different thing happen in circumstances that we are about to experience. Unless we can find a way to make these possible Laws of Nature less plausible than their blue-like cousins, this objection gets us nowhere.

Statistics Objection: Suppose that our life really is just a random collection of experiences. Then shouldn't we expect that a random sample from that life be similar to the proportions of the whole—so if almost all of what we have experienced so far conforms to induction, then probably what we experience in the future will too?

Response: If you examine the math closely enough, you will see that this doesn't work.

Section VI: The Major Objection: Infinity, Occam's Razor, and the Multiverse Distinction

In *On the Plurality of Worlds*, David Lewis considers a problem for his theory of modal realism which bears some similarity to ours—namely, that there would seem to be far more worlds in which we are deceived about the external world (brain in vat scenarios, for instance) than ones in which we are not. This probabilistic argument, it seems, should lead a modal realist to extreme forms of skepticism. To this problem, Lewis responds that it is impossible to make meaningful probabilistic assessments across the set of all possible worlds because of the infinite values involved. While we might, for example, be of the intuition that the positive integers outnumber the prime numbers and should be more probable if a number is chosen at random, but simple rearrangements of the number line can lead to the opposite intuition (115).

Translated to our argument, this objection amounts to the statement that it isn't fair to say that there are actually more chaos worlds than sane worlds in a mathematical (and thence probabilistic) sense. Along a similar line, it seems to be mathematically contradictory to weight the likelihood of each of an infinite set of items equally, as weighting an individual possibility with any non-zero value will result in a sum of all probabilities of infinity—not one.

Our response is in short that by invoking mathematics in this manner the multiverse theorist is jumping out of the frying pan and into the fire: It is mathematically impossible to assign equal weight to an infinite number of possibilities such that they sum to one. Yet this is precisely what multiverse theory seems to say must happen. There are three resolutions to this problem that we can see:

(1) Multiverse theory leads to an untenable conclusion, and thus is false.

(2) There is a problem with our argument that we should weight the worlds equally.

(3) There is in fact a meaningful and relevant way to give equal probabilistic value to each of an infinite number of possibilities.

If (1) is the case, then we have no need of further argument. If (2) is the case, it remains to be shown in what way our argument is unsound. Three seems the least tenable, but if it is true, it would seem to work in favor of our overall argument both because the multiverse theorist would have all the more reason to weight each world equally, and because it would suggest that probabilistic assessments across infinite sets are meaningful. Three would effectively take us back to our original argument in section IV, that we should weight the possibilities equally and that there are at least as many possibilities of chaos as of sanity. Thus in order for a multiverse theorist to build an objection to our argument out of (3), he would not only have to find a meaningful and relevant way to give equal probabilistic value to each of an infinite number of possibilities, he would have to find a way such that, e.g., the sum of the values of all the possibilities in which the cell phone drops is vastly greater than the sum of the values of all other possibilities.

A multiverse theorist might counter that this problem applies equally to non-multiverse theories, since they also typically propose an infinite number of possible futures which cannot be weighted equally as our argument says they should.[4] However, non-multiverse theories have a much easier job of holding that (2) is the case, because a non-multiverse theory can invoke this distinction between 'self-locating' possibility and 'metaphysical' possibility.

When a multiverse theorist like Lewis uses words like 'actual' or 'possible', he understands them to be quantifiers with a limited scope, meaning these things only with respect to the world in which they are spoken. When the multiverse theorist asks "Is the world sane or chaotic?" he means merely "Am I in one of the worlds that is sane, or one of the worlds that is chaotic?" Both answers posit the same number of existent entities; both

[4] Lewis states that he believes it is "just as reasonable for a modal realist as for anyone else to believe a priori that the actual world is clean." Presumably he would say the same for the belief that the actual world is sane. He does not explain why, however (Plurality, 121).

posit the same manner of existence for them. The multiverse theorist is choosing between theories that agree completely on all the non-indexical, non-self-locating facts. He is doing epistemology, not metaphysics. And epistemically speaking, possibilities that are both metaphysically identical and subjectively/empirically/phenomenologically indistinguishable (possibilities that Elga would call similar) ought to be weighted equally.[5]

In contrast, when the non-multiverse theorist asks "Is the world sane or chaotic?" he is asking about the way things are in the unrestricted, non-indexical, non-self-locating sense. For a non-multiverse theorist, the different answers to this question involve opposing metaphysical claims. When the non-multiverse theorist asks "Is the world chaotic or sane?" he means (among other things), "Does metaphysics suggest that there exists a chaotic universe, or a sane one?" The non-multiverse theorist is doing metaphysics, not epistemology, when he asks this question. It is at least in principle possible for the answer to these sorts of questions to be discoverable a priori; thus the non-multiverse theorist leaves himself open to the possibility of weighting some possible worlds more than others.

This is not to say that the non-multiverse theorist is going to be able to deduce a priori whether the world is chaos or sane. It is merely to say that the a priori route is potentially open to the non-multiverse theorist in a way that it is not for the multiverse theorist. For the multiverse theorist, the determination of how to distribute epistemic weight between these types of possibilities must be exclusively a posteriori. And since the a posteriori evidence for being in a chaotic world equals the a posteriori evidence for being in a sane world, reason compels the multiverse theorist to weight them the same.

Even if the multiverse theorist could surmount the above problem, he or she still has significant challenges to face when it comes to deciding in what manner the possibilities should be weighted differently. As an illustration, we will talk about the project of weighting the worlds by simplicity—Occam's Razor. Many of the problems we are about to present are equally applicable to any non-equal weighting of possible worlds, but we have chosen to apply them to Occam's Razor because of the special ontological and/or epistemological status that many people ascribe to simplicity.

For most people (non-multiverse theorists), Occam's Razor is formulated something like "The best theory is the simplest." For a multiverse theorist, the theories corresponding to the different possibilities—am I in world A or B, etc.—agree on all the objective facts; they are, it seems, of

[5]For a good argument explaining why, and an explanation of "similarity", see Elga, "Defeating Dr. Evil."

equal complexity. Thus for a multiverse theorist Occam's Razor must be formulated something like this in order to work: "The best theory is the one that involves me being in the simplest world."[6] Instead of measuring the simplicity of all that is, we are measuring the simplicity of a restricted sub-region of reality. This sort of restriction seems quite arbitrary.

The second is that the simplicity of a theory seems to be relative to the language of expression. Even rigorous formulations and definitions of simplicity—such as Kolgomorov complexity—work only once a language has been selected (Schmidhuber 28). It would seem, then, that our judgments about what is simple or complex are entirely dependent upon something very much contingent, as there is obviously no reason to preference one language over another. If the goal is to provide a non-arbitrary, a priori means of ranking worlds, then simplicity does not appear to be an adequate measure, however exactly simplicity is to be understood.

Furthermore, Schmidhuber and Solomonoff have suggested that an objective measure of simplicity might actually favor an infinitely large world containing every possible combination of laws and physical constants. Such a world would be much easier to formally specify than any particular world with its own particular set of laws, constants, and beings. This is because if we order worlds by the size of the 'program' that computes them, the program that computes all possible programs is actually quite short and simple (Solomonoff 598). It follows that an appeal to simplicity might leave the multiverse theorist in just as unhappy a situation.

Finally, there is a problem applicable to any way in which we order worlds. It is not enough only to order them objectively and say that we are more likely to be in worlds closer to the beginning of the list. We must have an account of how much more likely these worlds are—in other words, what the probability distribution across possible worlds actually is. This is a problem, because there are an infinite number of probability distributions that converge to one—and so a reason must be given to prefer one distribution (e.g. 1/(2^n), ½, ¼, 1/2, 1/4, 1/8, etc.) over another (e.g. 9/(10^n), 9/10, 9/100, 9/1000, etc.). Any choice of one over another would be entirely

[6]Note that a multiverse theorist will not be able to find any non-indexical proposition equivalent to the notion of "me" whereas most non-multiverse theorists are able to say something like "the person who occupies place *X* at time *T*." The multiverse theorist, in order to be able to do this, would need to specify what world he is in, and he does not have that information. This leads to interesting questions when it comes to analyzing the meaning of indexical statements in multiverse frameworks.

arbitrary, and therefore could hardly be a part of the a priori justification that the multiverse theorist is looking for.[7]

In conclusion: By assigning equal epistemic weight to every phenomenological counterpart, the multiverse theorist risks contradiction when thinking about the infinite array of such counterparts that he could be. Even setting aside these mathematical dangers, the multiverse theorist is confronted with the fact that there are at least as many chaotic worlds or futures as sane ones—and thus it seems he should expect his future to be contrary to the expectations of induction.

[7]Tegmark mentions an intriguing method of weighting and ordering the worlds that might get around this problem. By his own admission, though, it is still plagued by arbitrariness of the language sort (Mathematical Universe).

Works Cited

We would like to offer special thanks to Professor Megan Sullivan at the University of Notre Dame for her suggestions and advice as our project has progressed.

Elga, Adam. TS. Princeton University. "Defeating Dr. Evil with Self-locating Belief" Web. <http://www.princeton.edu/~adame/papers/drevil/drevil.pdf>.

——."Self-Locating Belief and the Sleeping Beauty Problem". Analysis 60.2, p. 144. April 2000.

Forest, Peter. "Occam's Razor and Possible Worlds." The Monist 65.4 (1982): 456–464.

Goodman, Nelson. "A New Riddle of Induction." Fact, Fiction, and Forecast. Cambridge, Harvard University Press, 1983. 59–83.

Hume, David. A*n Enquiry concerning Human Understanding*. davidhume.org, n.d., Web.

Lewis, David. "Causation." Philosophical Papers Volume 2. New York: Oxford University Press, 1986. 159–171.

——. *On the Plurality of Worlds*. United Kingdom: Wiley Blackwell, 2001.

——. "Sleeping Beauty: Reply to Elga". Analysis 61.3, p. 172. July 2001

Nozick, Robert. *Philosophical Explanations.* Cambridge: Harvard University Press, 1981.

Quine, W.V.O. "Natural Kinds." Ontological Relativity and Other Essays. New York: Columbia University Press, 1969. 114–138.

Schmidhuber, Jürgen. "Algorithmic Theories of Everything." N.P., n.d. Web. 17 Dec. 2012.

Solomonoff, Ray, J. "The Universal Distribution and Machine Learning." The Computer Journal 46.6 (2003): 598–601.

Tegmark, Max. "Is 'the Theory of Everything' Merely the Ultimate Ensemble Theory?" Annals of Physics 270.1 (1998): 1–51.

——. "The Mathematical Universe," Foundations of Physics, 8 October 2007: 23.