Modal Realism and Anthropic Reasoning

Mario Gómez-Torrente
Instituto de Investigaciones Filosóficas
Universidad Nacional Autónoma de México (UNAM)
mariogt@unam.mx

Abstract: Some arguments against David Lewis's modal realism seek to exploit apparent inconsistencies between it and anthropic reasoning. A recent argument, in particular, seeks to exploit an inconsistency between modal realism and typicality anthropic premises, premises common in the literature on physical multiverses, to the effect that observers who are like human observers in certain respects must be typical in the relevant multiverse. Here I argue that typicality premises are not applicable to the description of Lewis's metaphysical multiverse, where the proportions of metaphysically possible observers possessing the pertinent properties can be independently established by metaphysical reasoning. However, other kinds of reasonable anthropic premises can be seen to be consistent with and usable within modal realism.

Keywords: Modal realism, Anthropic reasoning, Multiverses, Observers, Typicality, Possible worlds

1. Introduction.

Following common usage, let us call "modal realism" David Lewis's view (notably in Lewis (1986)) that all metaphysically possible worlds possess the same sort of reality as the actual world.[1] By "anthropic reasoning", generalizing in a very wide but now common way the usage inaugurated by the physicist Brandon Carter (1974), I will understand reasoning that uses some ("anthropic") premise stating that certain theses d_1, d_2,… of a descriptive kind must stand

---

[1] As Lewis himself noted, the name unintendedly suggests the relatively weak view that there are real modal facts, and a more suitable title for the Lewisian view would be "possible worlds realism", but the usage is indeed common.

in a certain relation to certain claims o_1, o_2,… about the observer making the reasoning or a class of observers to which she belongs.[2] Thus, for example, one anthropic premise (roughly equivalent to Carter's "weak anthropic principle") says that the description of our (of us humans) local environment in the universe has to be consistent with the claim that observers exactly like us exist. Another anthropic premise (roughly equivalent to Carter's "strong anthropic principle") says that the physical laws and global description of our universe have to be consistent with the claim that observers exactly like us exist. Though heuristically important, Carter's principles are basically platitudes, even if, as is commonly done in the physical context, one takes the "strong anthropic principle" to presuppose the non-trivial thesis that there are other universes besides ours whose physical laws and global description are not consistent with the claim that observers exactly like us exist in them. But it is fairly common now to use the label "anthropic reasoning" in other cases where the relevant premise is not so clearly true and in essence requires (somewhat in opposition to the spirit of Carter's principles) that observers who are like the reasoning observer in some respect must be unremarkable in some sense. This idea may be cashed out in different ways. A structure for a class of anthropic premises inspired by this idea, which we might call "absolute typicality anthropic schema", and which gives the form of some more specific premises we will be concerned with in this paper, has it that a true description of a *multiverse* (or class of universes) M to which our universe belongs ought to imply the claim that observers who are like us humans in some respect R are *typical* in the universes of M. Though there is no fully agreed upon sense of "typical" in these contexts, the rough idea is that observers of a certain kind are to be considered typical in a multiverse when they constitute a substantive majority of all observers existing in universes of that multiverse. A key complicating factor here is that the number of all observers and the numbers of observers of pertinent kinds in the relevant multiverses is often infinite (and generally of the same infinite cardinality), and some special

---

[2] As Carter himself noted, the name unintendedly suggests that the relevant sort of reasoning involves only human observers, but it is intended to apply to reasoning involving observers in some (vague) general sense.

assumptions about how to compare the infinite sets involved is required in order to apply typicality anthropic premises; we will revisit this topic later.

A recent argument by Simon Friederich (2021), which echoes arguments previously put forward in print by Peter Forrest (1982) and Holly Thomas (1993) (and less formally by others[3]), uses anthropic reasoning, and specifically some instances of the absolute typicality anthropic schema, against Lewisian modal realism. As Friederich says, we can reasonably see modal realism as a multiverse theory, a theory that postulates the real existence of the universes in a certain huge class. There are several such theories currently under scrutiny in fundamental physics and cosmology, which postulate the existence of classes of universes, always much less numerous and much less unwieldy than the class of metaphysically possible worlds of modal realism, but huge, hard to manage classes nonetheless. The inherent difficulty of finding direct experimental tests for most such physical theories has led some physicists (and some philosophers) to seek indirect tests for them in anthropic reasoning, specifically in premises that use the idea of typicality. Thus, for example, a common way of reasoning in some of the literature on the so-called string theory "landscape" (a kind of multiverse arising from a combination of inflationary cosmology and string theory) is this: suppose that, using a certain package T of assumptions about how to measure and compare numbers of observers in the universes of the landscape, it can be shown that a substantive majority of physical observers will be like us in that they will inhabit universes where the value of the cosmological constant is in some way close to its value in our universe; then this confirms T, in the sense that, *ceteris paribus*, T will be taken to be more likely to be true than a different package T' that implies that a smaller proportion of observers in the universes of the landscape will be like us in that way.[4] Underlying this way of reasoning are implicit applications of another schema giving rise to anthropic premises: given two

---

[3] David Lewis (1986, 116) mentions George Schlesinger, Robert M. Adams, and J. J. C. Smart.

[4] Prominent items in this literature include Martel, Shapiro and Weinberg (1998), Bousso, Harnik, Kribs, and Perez (2007), and De Simone, Guth, Salem and Vilenkin (2008).

descriptions of a multiverse M, *ceteris paribus* the one of them that implies that there is a greater proportion of observers who are like us in some respect R in the universes of M is more likely than the other; we may call this the "comparative typicality schema".[5] Apparently taking his inspiration from such applications of the idea of typicality in physics, Friederich proposes that certain absolute typicality premises can and ought to be applied to modal realism, and that modal realism will clearly fail them. For example, according to him we can assume that if modal realism were true, it ought definitely to imply that observers who live in worlds where inductive practices like ours are reliable are typical in the class of possible worlds; but it is reasonably clear that, on the contrary, modal realism implies that such observers are not thus typical, so this provides us with a decisive test against modal realism.

My first main aim in this paper is to show that, whatever our antecedent views about the plausibility of modal realism, anthropic reasoning based on typicality premises does not provide a good argument against it. In Section 2 I will explain Friederich's arguments in more detail, and connect them with the earlier arguments by Forrest and Thomas. In Section 3 I will first recall certain facts about presumable cardinalities in the class of possible worlds that were arguably the basis for Lewis's rejection of Forrest's argument (and would probably have been the basis for his rejection

---

[5] It may be worth noting that the absolute and comparative typicality premises are strong premises independent of principles often taken as abstract codifications of anthropic reasoning. Thus, for example, they are independent of and stronger than Nick Bostrom's "Self-Sampling Assumption"—"One should reason as if one were a random sample from the set of all observers in one's reference class" (Bostrom (2002), 57). Unlike the relatively weak "Self-Sampling Assumption", typicality premises require that the proportion of observers who are like us in a certain respect (when the reference class is taken to be that of all observers existing in the universes of the relevant multiverse) has the special characteristic of being a substantive majority proportion. One of the main points below will be that this characteristic may obtain for the reference classes relevant in some physical contexts and not obtain for the reference class of all metaphysically possible observers. In some form or other, the Self-Sampling Assumption appears reasonable, and seems so to Bostrom. For brief remarks on a more problematic principle giving rise to anthropic premises, which Bostrom has called the "Self-Indication Assumption", see below.

of Thomas's argument); and secondly, I will argue that while these cardinality facts are less of a direct problem for Friederich's arguments than for Forrest's and Thomas's, they strongly suggest that the typicality anthropic premises of Friederich's arguments cannot be taken as reasonable in the metaphysical context where modal realism is put forward, regardless of whether premises of this kind can be assumed in physical contexts. In the final Section 4, I will suggest that though no typicality anthropic premises can be taken as reasonable in the metaphysical context, some related but crucially different anthropic premises might be used in anthropic reasonings within the framework of modal realism, but they would not be usable in a refutation of it.

2. Attempted anthropic refutations of modal realism.

Friederich's arguments can be reconstructed as using two instances of the absolute typicality anthropic schema. One instance, which we may call "Inductive Typicality", says that the true description of the class of possible worlds (to which the actual world of course belongs) has to imply the claim that observers who live in worlds where inductive practices like ours are reliable are typical in the class of possible worlds. The other instance, which we may call "Memorial Typicality", says that the true description of the class of possible worlds has to imply the claim that observers who live in worlds where apparent memories like ours are reliable are typical in the class of possible worlds. The argument that uses Inductive Typicality proceeds by assuming, certainly reasonably, that the true description of the space of metaphysically possible worlds implies that there is a huge number of worlds where observers follow inductive practices like ours, but these practices are unreliable: the future for those observers does not resemble the past in the ways we would expect. Friederich insists that there is no reason, and in particular none given by Lewis, to think that these observers are a more or less insignificant minority. In fact, as we will see (though this is not mentioned by Friederich), in a Lewisian framework there are good reasons to think that there are just as many observer-containing worlds where our inductive practices are unreliable as

there are observer-containing worlds where our inductive practices are reliable, and so it's reasonable to think that there are just as many observers of one kind as of the other. But then observers who live in worlds where inductive practices like ours are reliable are not typical in the class of possible worlds (even if they need not be atypical). Friederich concludes that this provides a decisive test against modal realism.

The argument that uses Memorial Typicality proceeds analogously, by assuming, certainly reasonably, that the true description of the space of metaphysically possible worlds implies that there is a huge number of worlds where observers have apparent memories like ours, but these apparent memories are unreliable: the past for those observers is not as those memories say it is; in fact, in many of these worlds, the observers with apparent memories like ours will have popped out into existence just a second earlier, with all their apparent memories in place, as in the case of the "Boltzmann brains" considered in fundamental physics. Friederich insists that there is no reason, and in particular none given by Lewis, to think that these observers are a more or less insignificant minority. (Again we will see that in a Lewisian framework there are good reasons to think that there are just as many observer-containing worlds where apparent memories are unreliable as there are observer-containing worlds where apparent memories are reliable, and so it's reasonable to think that there are just as many observers of one kind as of the other.) But then observers who live in worlds where apparent memories like ours are reliable are not typical in the class of possible worlds (even if they need not be atypical). The conclusion against modal realism again follows.

As we said, Forrest's and Thomas's arguments are related to Friederich's, but they use different premises about the relevant cardinality facts. While Friederich uses the presumable fact that possible observers living in inductively unreliable worlds or in memorially unreliable worlds are not an insignificant minority among possible observers, Thomas merely appeals to the presumable, weaker, fact that inductively unreliable worlds containing observers are "many"; and Forrest uses a premise (of dubious relevance, as we will see) according to which each world

containing observers, and in which reasonable abductive rules like Occam's razor generally hold, can be assigned a different class of infinitely many observer-containing worlds in which those rules fail to hold.

Thomas argues essentially as follows: "If modal realism is true, there are obviously *many* possible worlds where observers follow our inductive practices but, presumably unlike us, do so unreliably. If so, given that, according to modal realism, these observers have the *same sort of reality* that we do, we must take seriously the possibility that we are not different from them in living in a world where such practices are reliable. So we cannot reasonably rely on our inductive practices. But we *can* reasonably rely on our inductive practices. Therefore, we can conclude that modal realism is false." Note that, as we advanced, only a weak, qualitative idea that *many* pathological worlds exist according to modal realism is used, and no claim is made about how the class of such worlds compares cardinality-wise with the class of non-pathological worlds.

Forrest does use more quantitative cardinality comparisons in his relevant premise about modal realism, but this is substantively different from Friederich's absolute typicality anthropic premises. Forrest argues essentially as follows: "We think our world does not contain weird 'epiphenomenal particles', subatomic particles caused to exist by familiar particles but, unlike these, unable to influence causally other stuff, including our minds. In fact, we *can* reasonably reject the existence of epiphenomenal particles in our world through an application of Occam's razor, or other reasonable abductive rules. However, according to modal realism, for each world which looks the way our world looks to us and does not contain epiphenomenal particles, there is a *multitude*, in fact at least an infinity, of worlds which look the way our world looks to us and do contain epiphenomenal particles (in different numbers in each world, for example). So, according to modal realism we *cannot* reasonably reject the existence of epiphenomenal particles in our world, and in fact we can reasonably infer that our world does in all probability contain epiphenomenal particles! From the contradiction we can conclude that modal realism is false." Note that the key premise here, that for each non-pathological world there are many pathological worlds that add epiphenomenal particles to it, may

intuitively suggest that the class of "pathological" worlds is bigger than the class of "non-pathological" worlds, but in fact is compatible with the claim that the two classes have the same cardinality; indeed, as we are about to note, it's reasonable to think that they do have the same cardinality.

3. Cardinality and typicality.

Both in (1973) and (1986) Lewis appeals to extensions and refinements of a certain basic picture of Quine's (1969) when he needs to consider matters of cardinality of sets of worlds. We will reason within this picture not just when we describe Lewis's reaction to Forrest, but also when we deal with his presumable reaction to Thomas and with our rebuttal of Friederich, so let's begin by describing the picture. Suppose first that all that mattered to the specification of the contingent fundamental facts in a world was the specification, for each coordinate (w, x, y, z) in four-dimensional spacetime, of whether a matter particle of a certain unique kind fills the coordinate. Presumably there are simple possible worlds of this sort—Quine calls them "Democritean" worlds. For a given kind of matter particle, and given that there are $2^{\aleph_0}$ coordinates, the cardinality of the set of Democritean worlds (which is clearly independent of the choice of coordinate system) is evidently $2^{2^{\aleph_0}}$, also called beth-two ($\beth_2$). Lewis thinks that this suggests that the set of possible worlds has cardinality at least beth-two. But he also says that beth-two "is the number of (...) worlds at any level of generality that seems to me clearly called for" (Lewis 1973, 90, footnote). By this he means that the ways of extending and generalizing the basic contingent structure of a Democritean world that seem to him "called for" don't take us beyond beth-two. This is because these ways merely add finite or at most continuum-sized non-denumerable structure to that basic structure. Lewis mentions, for example, the need to add a metric to the four-dimensional manifold so as to distinguish, e.g., a Euclidean from a non-Euclidean spacetime, the possibility that the coordinates might be occupied not by points but by (still finitely specifiable) fields, and the possibility that there might be $2^{\aleph_0}$ kinds of matter or densities of

kinds of matter. Set-theoretic cardinal arithmetic tells us that these extensions and generalizations of the Democritean structure don't take us beyond beth-two. Other "un-Lewisian" but conceivable extensions presumably wouldn't take us beyond beth-two either, such as the possibility that space or time coordinates might be filled (also) by (non-material) mental elements of different kinds or by other kinds of material or non-material elements such as additional physical dimensions, or the possibility that the coordinate system were not a spacetime system, but just a system of placeholders for certain non-material elements.

To be sure, this idea about the exact cardinality of the set of possible worlds being beth-two can be disputed in many ways. To give just two examples, it might be objected that it is conceivable, or at least not clearly inconceivable, that some possible spacetime manifolds are more than continuum-many-dimensional, or that there are more than continuum-many possible kinds of matter. This might well be so—I doubt, and probably Lewis doubted too—that our epistemic capacities can get decisive answers to these modal matters. However, what seems a much less disputable assumption is that the contingent fundamental structure for a possible world must be given by some extension of the Democritean basic structure. And if we make this hard-to-dispute assumption, it is reasonable to think that we will be able to argue for conclusions about cardinality comparisons between sets of worlds that are in essence independent of the exact cardinality of the set of all worlds. Let's assume, then, that the contingent structure for a possible world is given by some extension of the Democritean basic structure, and, without real loss of generality as regards the results we will be reaching (as we will see), that the cardinality of the set of possible worlds is beth-two.

One of the conclusions that ensue is, for example, that the set of worlds containing epiphenomenal particles must have the same cardinality as the set of worlds not containing epiphenomenal particles. One problem in assessing this issue is that perhaps one same filling of the relevant extension of the Democritean structure could have alternative incompatible high-level descriptions, on one of which epiphenomenal particles are said to exist and on another of which they are said not to exist. But let's suppose, in a Humean-Lewisian fashion, that once we have a filling, it is determined (it

supervenes on the filling) whether there are epiphenomenal particles in it. Now it seems clear that the set of worlds without epiphenomenal particles will have cardinality beth-two. To see this, just consider the basic Democritean structure and begin to construct a filling supposing that there is complete emptiness before time t, that there is a particle p at a certain coordinate (w, x, y, t), and particles also at all remaining points with time coordinate t. These are our initial filling instructions. Now the set of sets of particles different from p in our construction has obviously cardinality beth-two. For every set S of this kind, define a completion C_S of our initial filling instructions by postulating that, on C_S, every particle q in S is pushed in some way by p at t along the direction that goes from p to q (we need not specify what result this produces after t in the distribution of particles, as long as the results are pairwise different for every S, as is clearly possible); and suppose further that on C_S, *every* particle aside from p pushes in some way p at t+1 along the direction that goes from q to p (again we need not specify the distribution of particles that this produces). This shows that the set of worlds without epiphenomenal particles has cardinality beth-two. But clearly the set of worlds with epiphenomenal particles will have at least cardinality beth-two and at most cardinality beth-two, so it will have cardinality beth-two as well. (Note that it would not be difficult to modify these arguments so as to run them for the case of Democritean structures that are more than continuum-many-dimensional, or where there are more than continuum-many possible kinds of matter; analogous things will happen with all the relevant arguments below, hence our remark above about the arguments establishing their conclusions without any loss of generality derived from the assumption that the cardinality of the set of possible worlds is beth-two.)

Another ensuing conclusion is that the set of (observer-containing) worlds where our inductive practices are reliable must have the same cardinality as the set of (observer-containing) worlds where they are not. To see this, it's enough to argue that the set of (observer-containing) worlds where our inductive practices are reliable must have cardinality beth-two. There are many conceivable ways to argue for this. A very simple one proceeds under the assumption that there is an observer-containing world (we believe

the actual world to be one such) where our inductive practices are reliable, there are no epiphenomenal particles, and there are (at some time or other) continuum-many empty (of matter) space coordinates; call one such world w. There are beth-two sets of empty space coordinates in such a world. For every such set S, let w_S be a world that is otherwise like w but fills exactly the coordinates in S at w with epiphenomenal particles (we need not specify by what normal particles they are produced, or how). The set of w_S worlds is clearly of cardinality beth-two, and in all of them our inductive practices are reliable. But also the set of worlds where our inductive practices are unreliable must have cardinality beth-two. A simple way to see this: think of the way w has been up until a certain time t and focus on a world w' which is like w up until t but where things continue in some unexpected way at t+1. Then for every set S of empty space coordinates in w', let w'_S be a world that is otherwise like w' but fills exactly the coordinates in S at w' with epiphenomenal particles (we need not specify by what normal particles they are produced, or how). The set of w'_S worlds is clearly of cardinality beth-two, and in all of them our inductive practices are unreliable.

Lewis bases his rejection of Forrest's argument on such presumable facts about cardinalities. Much as it's true that, as Forrest said, for each non-pathological world there are many pathological worlds that add epiphenomenal particles to it, careful consideration of the situation shows that the relevant cardinalities are the same, and that that truth is irrelevant for Forrest's purposes. (It is equally true, and equally irrelevant, that for each pathological world there are many non-pathological worlds.) So, Forrest's claim that, according to modal realism, in all probability we live in a world with epiphenomenal particles, cannot be grounded in a comparison of cardinalities between the complementary sets of worlds with and without epiphenomenal particles.

As Lewis notes, however, a claim like Forrest's might conceivably be grounded in some special assumption about how to compare the relevant infinite sets that didn't simply rely on cardinalities—in the same way that, as we noted above, special assumptions about how to measure and compare relevant infinite sets of universes and observers are used and evaluated in the

literature on physical multiverses. Forrest's claim, says Lewis, might perhaps appeal to some argument that relied on "taking limits" over some suitable ordering of the possible worlds, for example. For an analogy, think that even though the cardinality of the set of prime numbers is the same as that of the set of non-prime numbers, one might reasonably argue that the set of primes is somehow of a smaller size than that of the non-primes by noting that their limiting relative frequency over the natural order of the natural numbers is 0. Perhaps the set of abductively non-pathological worlds could be argued to be smaller than the set of abductively pathological worlds in this way. But Lewis also says, quite reasonably, that there is just no known conception of a natural (well-)order of the possible worlds, and so this avenue for the rescue of Forrest's argument is closed. Lewis doesn't consider other possibilities. There is the possibility that some natural measure function could be defined on the subsets of the set of possible worlds without exploiting a well-order, but in such a way that the set of non-pathological worlds came to have a small measure and the set of pathological worlds came to have a large measure. Here the analogy would be with the measurable subsets of an n-dimensional Euclidean space and the Lebesgue measure over them. Another possibility would be that some finite set of worlds could be selected of which we had reason to think that it was representative of the proportion of non-pathological worlds vs. pathological worlds, and that it turned out that the pathological worlds in this set were many more than the non-pathological worlds. But again we can reasonably deny that there is any expectation that a measure function or a representative finite set with all those properties exist. In the unwieldy metaphysical multiverse, as we have already begun to experience, all possibilities are on a par, so to speak.

Importantly, the situation with the space of metaphysically possible worlds contrasts sharply with the situation in fundamental physics and cosmology. A physical multiverse theory always postulates a comparatively much less abundant class of universes than in modal realism, a class of universes where many physical laws and assumptions are taken to hold, and which always emerge through some process of physical generation. Thus, for example, the universes in the string theory landscape emerge via a recursive

operation of the mechanisms of inflationary expansion and quantum tunneling starting from the assumed singularity known as the Big Bang. This implies that there must be a physically natural temporal partial ordering with minimal elements of the universes in the landscape, and suggests ideas about how to define measures on the subsets of this ordering. (For example, (1) choosing (essentially) a given time after the singularity and trying to calculate the number of the relevant sets of universes restricted to those appearing before that time (see e.g. Linde, Linde and Mezhlumian (1994)); or (2) doing that together with the stipulation of some factor that gives more weight to universes as these are produced along the temporal line (see e.g. De Simone, Guth, Salem and Vilenkin (2008)); another option is (3) choosing a so-called causal diamond (i.e. a geodesic along a timeline starting from the Big Bang and up to a time when no physical interactions occur any longer, along with the volume around the geodesic that can interact causally with the geodesic), which will be finite, and calculating the numbers of the relevant sets of observers as restricted to those in the diamond (see e.g. Bousso, Harnik, Kribs, and Perez (2007)); and others.) But these routes to the definition of orderings and measures are closed in the case of the hard-to-manage class of metaphysically possible worlds, which have no comparable unifying principles or processes of generation governing them.

How does Lewis go from his rejection of Forrest's argument to (a justification for) his acceptance of Occam's razor and other reasonable abductive rules? Unfortunately, he is not very explicit about this. Very clearly Lewis implies that the modal realist ought to accept the existence of epiphenomenal particles in the actual world if the cardinality of the set of worlds that contain them is overwhelmingly greater than that of the set of worlds not containing them. And then he goes on to note that this is not the case, and in fact the odds are even, so to speak. I think it seems reasonable, then, to see Lewis as thinking that, absent any pressure to view the existence of epiphenomenal particles as somehow likely, and given that believing in them and thus rejecting established abductive rules would introduce a considerable lack of equilibrium within our set of beliefs and epistemic practices, we can with "confident hope" (Lewis (1986), 117) choose to keep our established abductive rules

and believe that we live in a world without epiphenomenal particles. This will preserve equilibrium with little or no alteration of our initial system of beliefs. In his famous remarks on philosophical method and the role of opinion in it toward the end of his Introduction to the first volume of his *Philosophical Papers* (Lewis (1983)), Lewis says that one task of the philosopher is that of settling on one of the equilibrated sets of commonsense plus theoretical beliefs that philosophy lays out before us, but also that which equilibrated set one chooses is a matter of opinion. Our case would probably be one of the cases to which Lewis would apply this conclusion, though it is also a case in which a reasonable conservative attitude shaping the relevant opinions will clearly settle for a set of beliefs that includes the belief that we live in a world without epiphenomenal particles.

Similarly, although Thomas's claim that the worlds where our inductive practices are unreliable are many is not refuted, but vindicated within Lewis's usual conception of the space of possible worlds, we know that the worlds where our practices are reliable are many as well, in fact just as many. So, there is no pressure for the modal realist to believe that we live in an inductively unreliable world, it is open to her to believe that we live in an inductively reliable world, and this is ultimately what Lewis chooses to do, presumably on account of the fact that in doing otherwise he would introduce a revolutionary imbalance in his system of beliefs.

We can't know for certain what Lewis's reaction to Friederich's arguments would have been, and it's unclear what it could be, because, unlike the Forrest and Thomas arguments, Friederich's arguments don't depend for their effectiveness on the idea that the relevant pathological worlds are much more abundant than the non-pathological worlds (or on special assumptions about how to compare relevant infinite sets of worlds or observers). Friederich's arguments employ absolute typicality anthropic premises to the effect that modal realism, if true, ought to imply the claim that observers living in non-pathological worlds are typical. And they reach its anti-modal-realist conclusion because presumably the true description of the metaphysical multiverse implies that observers living in non-pathological worlds are neither typical nor atypical—

there are just as many observers of this kind as observers living in pathological worlds.

That this is indeed a presumable implication of the true description of the space of possible worlds is easy to see in view of our earlier results. We can use one of our earlier proofs in order to argue again that the set of observer-containing worlds where our inductive practices (/memories) are reliable must have cardinality beth-two. But this clearly implies that the set of possible observers living in worlds where our inductive practices (/memories) are reliable must have cardinality beth-two as well. And also the set of possible observers living in worlds where our inductive practices (/memories) are unreliable must have cardinality beth-two, given our earlier proof that the set of observer-containing possible worlds where our inductive practices are unreliable must have cardinality beth-two. This adapts easily also to a proof that the set of observer-containing possible worlds where memories like ours are unreliable must have cardinality beth-two. So again this clearly implies that the set of possible observers living in worlds where our inductive practices (/memories) are unreliable must have cardinality beth-two as well.

In fact, it is reasonable to think that when both P and not-P are qualitative, consistent properties an observer may have, both the set of metaphysically possible observers having P and the set of metaphysically possible observers having not-P will usually have beth-two cardinality. For let w be a world where an observer O has P (not-P); usually we will be able to modify in beth-two ways the Democritean or extended-Democritean structure surrounding O in w by using our modal imagination, while retaining O and its possession of P (not-P) in the modifications, as we did in the proofs above. It seems it can be said that the normal situation when P and not-P are qualitative, consistent properties of observers, will be that there will be beth-two possible observers having P and beth-two possible observers having not-P.

But now, given how reasonable all this is, the casual adoption of typicality anthropic premises in the context of a discussion of modal realism turns out to be thoroughly implausible. As we saw, these premises (in the form of comparative typicality premises) were in essence originally designed to provide a statistical confirmation

test for a context (the physical context) where many multiverse-theory + measure-and-compare-assumptions packages are in principle viable which imply different cardinality proportions for the sets of observers having P and not-P in the assumed multiverse. The situation with physical multiverse theories is as a consequence epistemically very different from the situation in a metaphysical context where the multiverse of modal realism is being considered. A theory of a physical multiverse always selects, from among the metaphysically possible universes, a comparatively very small set of universes (occasionally a finite set, but usually a countably infinite set) where a good deal of established and hypothetical physical postulates are required to hold. As a result, the usual, presumably *a priori* constraint in the class of possible worlds, that the observers in these universes who have a certain property P and the observers who have not-P have the same cardinality (and cannot be differentiated by natural measures), is automatically lifted. And it is normally a comparatively non-obvious and often difficult matter whether that good deal of physical postulates imply that the observers in those universes that have P will be typical. In fact, these theoretical packages are proposed in part precisely with the expectation or aim that they will imply that observers who are like us in certain respects are typical in the universes of the relevant multiverse, and that this will provide confirmation of the theory as compared with other theories. In this context, if we are choosing between two packages which are otherwise similar as to their implications, it is certainly reasonable to choose that which implies that observers having a property P that we (believe to) have ourselves are statistically more typical in its corresponding multiverse. (Though even in this context the "loser" package isn't thereby thoroughly refuted.) In this epistemic situation, it does make sense to adopt a typicality anthropic premise. But it should by now be clear that the same cannot be said of the situation of the proponent of modal realism.

The metaphysical context where modal realism is proposed is indeed epistemically quite different from a context where a physical multiverse is being considered. In the metaphysical context, although there is a dispute about whether non-actual possible worlds exist in the same sense as the actual world, there isn't a

substantial variety of views concerning the cardinality proportions of relevant sets of possible worlds and observers. Usually *we already have a strong belief*, and certainly the Lewisians have a strong belief, that the set of possible observers having P and the set of possible observers having not-P have the same cardinality. (And we presume, with good reason, that there are no reasonable special assumptions about how to compare relevant infinite sets of observers leading to the conclusion that the P-observers are somehow less frequent than the not-P-observers). But if so, it is simply inappropriate to adopt a typicality anthropic premise concerning P, designed for a different context, where we have essentially no reason not to think of ourselves as typical in usual respects. For in the metaphysical context we do have strong antecedent reasons to think of ourselves as neither typical nor atypical.

Quite generally, statistical premises are appropriate in contexts where a certain relevant kind of knowledge is lacking. If we don't know whether a certain person *a* has a certain property P, but know that a large percentage of people with some properties we know that *a* has do have P, we are reasonably entitled to believe, or assign a high probability to the claim that, *a* has P; but of course, if we then find out that *a* has not-P, we are no longer entitled to that belief, and the statistical touchstone loses its applicability. Though different, the case that concerns us is relevantly similar. If we don't know the true theoretical package for a certain postulated physical multiverse, including the proportion of observers having P in that multiverse, *ceteris paribus* we are reasonably entitled to believe, or assign a high probability to, a package that implies that P (which we believe to have ourselves) is had by a substantive majority of observers. If we didn't know the true theory of the metaphysical multiverse concerning the proportion of observers having P, *ceteris paribus* we would be reasonably entitled to believe, or assign a high probability to, a theory that implied that P (which we believe to have ourselves) is had by a substantive majority of possible observers. But more than usually we have strong reasons to believe that we know the proportion of observers having P in the metaphysical multiverse, and that this proportion is such that the observers having P are not typical (nor are the observers having not-P). Given this fact, the application of that statistical confirmation test loses its relevance. If

via some surprising kind of fully convincing reasoning we reached the conclusion that we knew the proportion of observers having P in a particular physical multiverse, then, even if the proportion of observers having P came out low or equal to that of observers having not-P, use of typicality premises to choose between theories of that multiverse would lose its relevance as well.

Related caveats and general observations are not infrequent in the literature on anthropic premises. Take for example much of the literature surrounding J. Richard Gott's (1993) controverted but very often accepted method for giving a reasonable estimate of the expected duration of a thing or process. (As is well-known, its most popular application has been to the formulation of estimates of the duration of human civilization.) When the desired degree of confidence is the scientific standard of 95%, Gott's proposal can be stated as follows (compare also the presentation in Gott (2001), ch. 5): If a reasoning observer is located at a random time in the duration of a certain thing or process T, there is a 95% chance the observer is in the middle 95% of the period in which T is observable, which implies that there is a 95% chance that the future duration of this period at the observer's time is between 1/39th and 39 times as long as its past. (1/39 is precisely half of 5%.) The schema "There is a 95% chance that the future duration of the period in which T is observable at the reasoning observer's time is between 1/39th and 39 times as long as its past" gives rise to anthropic premises in our sense.

But note that Gott's method relies on assuming that certain observers' position (normally, ours in the present time) with respect to the relevant duration is random. Many criticisms launched at the method involve exhibiting examples where it gives wrong results, but where there is antecedent reason to think of the chosen observers' position as peculiar and non-random in some relevant respect. However, Gott has always been reasonably careful to note that the application of his method requires a previous determination that one has no reason to think of the chosen observers' position as non-random in relevant respects (see Monton and Kierland (2006) for philosophical discussion sympathetic to Gott). More generally, we can say here that, in order to apply anthropic premises based on assuming ourselves to be random, or typical, or to possess some

other statistical characteristic, the reasoner must not have intuitively sufficient reasons to think that we are not random, typical, etc. in the relevant respects. Anthropic reasonings based on such premises are essentially probabilistic or statistical arguments, and there is no necessity in the assumption that the relevant observers are random, typical, etc. in the relevant ways. If some inferences from accepted data prove or suggest in a sufficiently strong way that they in fact are not, we just cannot apply the relevant anthropic premises. What we have argued above is that this caveat is not heeded by Friederich when he argues against modal realism on the basis of typicality anthropic premises.[6]

Does this mean that anthropic reasoning has no role to play for a proponent of modal realism? I think not, and I will sketchily explain some reasons why in the next and final section.

4. *Anthropic reasoning for modal realists.*

Are there anthropic premises that the modal realist ought to or at least could consistently, and reasonably, be committed to? Well, we have in fact already come across a certain form of anthropic claims that will usually deliver reasonable anthropic premises, even for the modal realist: a description of a class of universes C to which our universe belongs has to be consistent with, or even imply, the claim that observers who are like us in some respect R are *not atypical* in the universes of C. We saw that Lewis himself admitted that a good argument to the effect that possible observers who (presumably like us) live in abductively reliable worlds are a tiny minority among all possible observers would put pressure on his modal realism. However, as we also saw, the presumable true description of the space of metaphysically possible worlds will, at least in the cases that have presented themselves for consideration, not only be consistent

---

[6] We should note that Friederich is of course aware that the attractiveness of a typicality assumption or a randomness assumption depends on the reasoning observer's background knowledge concerning her position with respect to the reference class of observers that she is considering in her reasoning (see Friederich (2021), 109ff.). However, he doesn't show any awareness that this dependence may vitiate his critique of modal realism.

with, but even imply claims of non-atypicality, and so modal realism will be consistent with and could in fact reasonably adopt the corresponding anthropic premises.

Also, Lewis (1986, 132) himself notes (without citing Carter) that something like Carter's strong anthropic principle can be used within the framework of modal realism in order to dissolve in a certain way a sense of surprise (if we have it) that we live in a habitable world, for example. This might indeed appear surprising, given that there are so many possible worlds, presumably beth-two once more, that are uninhabitable. Our universe's description has to be consistent with the truth that we exist, says Carter's strong anthropic principle. This immediately implies that our world's true description cannot imply that our world is uninhabitable, for we do exist and so inhabit it. It's no surprise that we live in a habitable world, for in an uninhabitable we would not exist (and would not be saying all these things). Modal realism is obviously consistent with a possible worlds version of the strong anthropic principle, as it has no counterintuitive implication concerning the true description of the actual world.

Likewise, a principle closely analogous to Carter's strong anthropic principle can be used in order to dissolve in a certain way a sense of surprise (if we have it) that, for example, we are epistemically responsible observers who use the forms of reasoning we use and develop discussions such as the present one led by those forms of reasoning (even if, conceivably, we might live in a world where these forms of reasoning are not always correct). This might indeed appear surprising, given that there are so many possible observers, presumably beth-two once more, who don't care about epistemic responsibility. That our world's description has to be consistent with the claim that we are epistemically responsible observers is a reasonable anthropic premise, since it is obvious that we are and that that's what leads us to these very discussions in the first place. This immediately implies that our world's true description cannot imply that our world does not contain epistemically responsible observers, for we regularly do things like responsibly considering questions like these and so are epistemically responsible. It's no surprise that we are epistemically responsible, for if we were irresponsible we would not be considering these very

questions, and we obviously are. Modal realism is clearly consistent with this anthropic principle, again because it has no counterintuitive implication concerning the true description of the actual world.

For the same reason, Lewisian modal realism is also consistent with applications of anthropic reasoning—indeed, of typicality anthropic premises—in fundamental physics and cosmology. The fact—if it is a fact—that we are typical observers in the "landscape" multiverse, say, along with all its potential implications, doesn't contradict any claim of modal realism. In particular, that fact is independent of modal realism's presumable implication that observers who are like us in usual respects are neither typical nor atypical in the metaphysical multiverse. To the extent that modal realism seeks, or assumes itself, to be continuous with our best scientific conception of the actual world, modal realists can rest content that this aim is not spoiled by their views on the class of metaphysically possible observers.

Another reasonable and potentially philosophically substantive anthropic schema, or principle for the construction of particular anthropic premises, that again involves the description of the actual world, and that there is no reason to think modal realism fails to be consistent with, is this: when there is no known reason to think of the reasoning observer as peculiar in a certain respect (e.g. as non-random as to her location along a certain dimension), a description of the actual world is likely to imply the claim that there is an abundant class of actual observers to which that observer belongs and who are like that observer in that respect. Call this the *actual mediocrity anthropic principle*.[7] Again we have here an underlying idea that an observer such as anyone of us must be unremarkable, at least in respects where there is no known reason to think otherwise

---

[7] The actual mediocrity anthropic principle is related to what Bostrom (2002) calls the "Self-Indication Assumption", but appears far more reasonable. The "Self-Indication Assumption" says that, of two descriptions (of a universe, multiverse, etc., of which the reasoning observer forms part), *ceteris paribus* the one of them that implies that there is a greater *number* of observers is more likely than the other (cf. Bostrom (2002), 66). There is indeed no intuitive reason why this ought to be a good principle, as made perhaps poignantly clear by Bostrom's well-known "Presumptuous Philosopher" thought experiment (Bostrom (2002), 124).

(specifically when it comes to comparing her location along a certain relevant dimension with that of other observers). As we would expect, this unremarkability is not unqualified—it is restricted to cases where there is no known reason to think of a given observer as peculiar (e.g. as non-random along a certain dimension). It is not excluded that in many cases, in particular in cases where there are compelling reasons to think of a particular observer as peculiar in a certain respect, she *is* peculiar or in that respect.

One very basic philosophical application of the actual mediocrity anthropic principle might be in a certain argument against solipsism. The principle implies that if there is no known reason to think of myself as peculiar as to the property of existing, a true description of the actual world is likely to imply that there is an abundant class of actual observers to which I belong and who are like me in that they exist. Assuming there is in fact no truly known reason to think of myself as peculiar as to the property of existing, the principle implies the specific anthropic premise that a true description of the actual world is likely to imply that there is an abundant class of actual observers to which I belong and who are like me in that they exist, and hence that there must exist many other observers besides myself in the actual world. A variation on this premise, that would also follow from the actual mediocrity principle under the assumption that there is no truly known reason to think of myself as peculiar in being a sentient body, would be the anthropic premise that a description of the actual world is likely to imply the claim that there is an abundant class of actual observers to which I belong and who are like me in that they are sentient bodies.

Many other respects as to which there would seem to be no known reasons to think of myself as peculiar can be plugged into the actual mediocrity anthropic principle, with intuitively appealing results. The principle implies that if there is no known reason to think of myself as peculiar as to the property of mentally representing a physical world with the apparent general characteristics of the actual world, a description of the actual world is likely to imply the claim that there is an abundant class of actual observers to which I belong and who are like me in that they mentally represent a physical world with the apparent general characteristics of the actual world. Assuming there is in fact no

known reason to think of myself as peculiar as to the property of mentally representing a physical world with the apparent general characteristics of the actual world, the principle implies the specific anthropic premise that a description of the actual world is likely to imply the claim that there is an abundant class of actual observers to which I belong and who are like me in that they mentally represent a physical world with the apparent general characteristics of the actual world, and hence that there must exist many other observers besides myself who represent a physical world roughly like the world I think I live in. This is an interesting conclusion, especially insofar as it runs against the skeptical hypothesis that whatever observers may exist in the actual world need not be related, or might exist without connection with one another. The present anthropic premise implies that at least many of these observers will probably be similar in an important respect concerning what their mental lives represent. In similar ways one might argue in an anthropic fashion that many actual observers are likely to share their phenomenal representation of the physical color spectrum with me, that many actual observers are likely to share their subjective representation of the passage of time with me, etc. These applications and the ones in the previous paragraph are quite obviously compatible with both the letter and the spirit of modal realism.

Anthropic premises of Gott's type also seem to be usable from the point of view of a Lewisian modal realist. Recall that in the standard case, and simplifying slightly, Gott's suggestion is that if we have no reason to think that a reasoning observer is not located at a random time in the duration of a certain thing or process T, there is a 95% chance that the future duration of T at the observer's time is between 1/39th and 39 times as long as its past. Now as Gott himself suggests, the idea behind his method can be applied to dimensions other than time and intervals other than durations. Generalizing in this direction, it is natural to postulate the following anthropic principle, to be applied under the usual proviso that there is no known reason to think that the reasoning observer is not random in the relevant respect: Let T be an arbitrary thing and I a bounded real interval naturally associated with T (its duration, for example); then if a reasoning observer is randomly located in I, there is a 95% chance that the right subinterval of I at the observer's

location is between 1/39th and 39 times as long as its left subinterval.

A potentially attractive philosophical application of this principle might exploit my presumable *epistemic* unremarkableness along a number of dimensions. For example, taking T to be the notion of knowledge practically available to human observers, there would appear to be no known reason to think of myself as non-random as to my location along the associated, presumably bounded, interval of *amounts of knowledge practically available to a human observer*. If so, our generalized Gottian principle will ultimately entail that there is a 95% chance that the amount of knowledge practically available to me is between 1/39 and 39/40 times the maximum amount of knowledge practically available to humans. This is interesting, insofar as it would suggest that the amount of knowledge not practically available to me but available to other present or future humans is relatively meager, or at any rate more meager than one might perhaps expect.[8] (Note that this result would obtain independently of the truth of the matter concerning the duration of the human race, in particular independently of whether the human race turned out to last a lot longer than usual—anthropic—doomsday arguments allow.) Similar anthropic reasonings would lead us to conclude, for example, that specifically *a priori* knowledge not practically available to me but available to other present or future humans is relatively meager, or that specifically metaphysical or broadly philosophical knowledge not practically available to me but available to other present or future humans is relatively meager. Anthropic conclusions of this sort are certainly not inconsistent with modal realism, and are not intrinsically unattractive: one might well think that they give a relatively concrete expression to a certain sense of modesty that could turn out to provide the right perspective on our epistemic

---

[8] Our generalized Gottian principle involves real bounded intervals and thus, in this particular application, that the maximum amount of knowledge practically available to humans is finite, which seems a reasonable assumption but might be disputed under some views. What the application of the principle implies in this application is, strictly put, that if the maximum amount of knowledge practically available to humans is finite, then it is also relatively meager.

position in the actual world (though presumably not in all possible worlds).

5. Conclusion.

To sum up: even though what we have called typicality anthropic premises are in many cases incompatible with Lewis's modal realism, this is not a problem for this doctrine, against what Friederich argues. Modal realism is a metaphysical theory with independently plausible and largely uncontroversial (for Lewisians, at the very least) implications concerning the (usually equal) proportions of worlds and observers possessing particular properties. These implications cannot be taken to have been refuted by premises designed for confirmation purposes in the physical context, where comparable or corresponding proportions are really unknown and much more controversial. However, other kinds of anthropic reasonings, including reasonings based on suitable metaphysical versions of Carter's strong anthropic principle, and on what we have called actual mediocrity premises and a generalized Gott principle, are not unreasonable forms of anthropic reasoning consistent with modal realism.

References

Bostrom, N. (2002), *Anthropic Bias. Observation Selection Effects in Science and Philosophy*, Routledge, London.

Bousso, R., R. Harnik, G. D. Kribs, and G. Perez (2007), "Predicting the Cosmological Constant from the Causal Entropic Principle", *Physical Review D* 76, 043513, 1-16.

Carter, B. (1974), "Large Number Coincidences and the Anthropic Principle in Cosmology", in M. S. Longair (ed.), *Confrontation of Cosmological Theories with Cosmological Data*, Reidel, Dordrecht, 291-298.

De Simone, A., A. H. Guth, M. P. Salem and A. Vilenkin (2008), "Predicting the Cosmological Constant with the Scale-Factor Cutoff Measure", *Physical Review D* 78, 063520, 1-16.

Forrest, P. (1982), "Occam's Razor and Possible Worlds", *The Monist* 65, 456-464.

Friederich, S. (2021), *Multiverse Theories. A Philosophical Perspective*, Cambridge University Press, Cambridge.

Gott, J. R. (1993), "Implications of the Copernican Principle for our Future Prospects", *Nature* 363, 315-319.

Gott, J. R. (2001), *Time Travel in Einstein's Universe. The Physical Possibilities of Travel through Time*, Houghton Mifflin, New York.

Lewis, D. K. (1973), *Counterfactuals*, Blackwell, Oxford.

Lewis, D. K. (1983), *Philosophical Papers, vol. I*, Oxford University Press, Oxford.

Lewis, D. K. (1986), *On the Plurality of Worlds*, Blackwell, Oxford.

Linde. A., D. Linde and A. Mezhlumian, "From the Big Bang Theory to the Theory of a Stationary Universe", *Physical Review D* 49, 1783-1826.

Martel, H., P. R. Shapiro and S. Weinberg (1998), "Likely Values of the Cosmological Constant", *The Astrophysical Journal* 492, 29-40.

Monton, B. and B. Kierland (2006), "How to Predict Future Duration from Present Age", *Philosophical Quarterly* 56, 16-38.

Quine, W. V. (1969), "Propositional Objects", in Quine, *Ontological Relativity*, Columbia University Press, New York, 147-155.

Thomas, H. G. (1993), "Modal Realism and Inductive Scepticism", *Noûs* 27, 331-354.