What Is a Physically Reasonable Space-Time?*

John Byron Manchak†‡

Cosmologists often use certain global properties to exclude “physically unreasonable” cosmological models from serious consideration. But, on what grounds should these properties be regarded as physically unreasonable if we cannot rule out, even with a robust type of inductive reasoning, the possibility of the properties obtaining in our own universe?

1. Introduction. Recent results show one sense in which the global structure of space-time cannot be fully established (Manchak 2009a). It seems that, excluding certain pathological examples, every cosmological model is empirically underdetermined; no amount of observational data we could ever (even in principle) accumulate can force one and only one cosmological model on us. Additionally, one can show that even under the assumption of an inductive principle—that the physical laws we determine locally are applicable throughout the universe—these general epistemological difficulties remain.

However, it may be that we are able to make partial determinations concerning important global properties. For example, it might be possible to discover whether our universe is causally well behaved in some respect. The first task of this article will be to demonstrate that for many global properties of interest, even this partial determination is impossible. This certainly exacerbates the epistemic plight of the cosmologist. But, in ad-

*Received September 2010; revised November 2010.
†To contact the author, please write to: Department of Philosophy, 361 Savery Hall, Seattle, WA 98195; e-mail: manchak@uw.edu.
‡I am grateful to Jeff Barrett, Bob Geroch, David Malament, John Norton, and Kyle Stanford for comments on an earlier draft. I also wish to thank audiences at Pittsburgh and Oxford for valuable discussions.
WHAT IS A PHYSICALLY REASONABLE SPACE-TIME? 411

dition, these underdetermination results seem to have implications for what counts as a “physically reasonable” cosmological model.

Cosmologists often use certain global properties (e.g., those relating to causal misbehavior) to exclude some models from serious consideration. But, on what grounds should these properties be regarded as physically unreasonable if we cannot rule out, even with a robust type of inductive reasoning, the possibility of the properties obtaining in our own universe? The second task of this article is to carefully investigate this question.

2. Background Structure. We start by reviewing the relevant background formalism of general relativity.1 A relativistic space-time is a pair of mathematical objects \((M, g_{ab})\), where \(M\) is a connected four-dimensional manifold (without boundary) that is smooth (infinitely differentiable). Here, \(g_{ab}\) is a smooth, nondegenerate, pseudo-Riemannian metric of Lorentz signature \((+, -, -, -)\) defined on \(M\). Each point in the manifold represents an “event” in space-time. For each point \(p\) in the manifold, the metric assigns a light cone structure in the tangent space \(M_p\). Any tangent vector \(\xi^a\) in \(M_p\) will be timelike (if \(g_{ab}\xi^a\xi^b > 0\)), null (if \(g_{ab}\xi^a\xi^b = 0\)), or spacelike (if \(g_{ab}\xi^a\xi^b < 0\)). Null vectors create the “cone” structure. Timelike vectors are inside the cone, while spacelike vectors are outside. A time-orientable space-time is one that has a continuous timelike vector field on \(M\). A time-orientable space-time allows us to distinguish between the future and the past lobes of the light cone. In what follows, we assume that space-times are time orientable.

For some interval \(I \subseteq R\), a smooth curve \(\gamma: I \to M\) is timelike if the tangent vector \(\xi^a\) at each point in \(\gamma[I]\) is timelike. Similarly, a curve is null (respectively, spacelike) if its tangent vector at each point is null (respectively, spacelike). A timelike curve is future directed if its tangent vector at each point lies in the future lobe of the light cone. For any two points \(p, q \in M\), \(q\) is to the timelike future of \(p\) (written \(p \ll q\)) if there exists a timelike, future-directed curve \(\gamma\) from \(p\) to \(q\). A future-directed curve from \(p\) to \(q\) that is either timelike or null indicates that \(q\) is to the causal future of \(p\) (written \(p < q\)). These relations allow us to define the following sets of points in \(M\): \(I^-(p) = \{q: q << p\}\), \(I^+(p) = \{q: p << q\}\), \(J^-(p) = \{q: q < p\}\), and \(J^+(p) = \{q: p < q\}\). The set \(I^-(p)\) will be used extensively throughout the article and is called the observational past of \(p\). This set represents those points that can possibly be observed from \(p\).2

A point \(p \in M\) is a future endpoint of a future-directed causal curve \(\gamma: I \to M\) if, for every neighborhood \(O\) of \(p\), there exists a point \(t_0 \in I\)

1. Details can be found in Hawking and Ellis (1973) and Wald (1984).
2. One uses the set \(I^+(p)\) instead of \(J^+(p)\) to represent the observational past of \(p\) for reasons of mathematical convenience.
such that $\gamma(t) \in O$ for all $t > t_0$. A past endpoint is defined similarly. For any set $S \subseteq M$, we define the past domain of dependence of $S$ (written $D^-(S)$) to be the set of points $p \in M$ such that every causal curve with past endpoint $p$ and no future endpoint intersects $S$. The future domain of dependence of $S$ (written $D^+(S)$) is defined analogously. The entire domain of dependence of $S$ (written $D(S)$) is just the set $D^-(S) \cup D^+(S)$.

A set $S \subseteq M$ is achronal if no two points in $S$ can be connected by a timelike curve. The edge of a closed, achronal set $S \subseteq M$ is the collection of points $p \in S$ such that every open neighborhood $O$ of $p$ contains a point $q \in I^+(p)$, a point $r \in I^-(p)$, and a timelike curve from $r$ to $q$ that does not intersect $S$. A set $S \subseteq M$ is a slice if it is closed, achronal, and without edge. A set $S \subseteq M$ is a spacelike hypersurface if $S$ is a three-dimensional submanifold such that every curve in $S$ is spacelike.

Two space-times $(M, g_{ab})$ and $(M', g'_{ab})$ are isometric if there is a diffeomorphism $\phi: M \to M'$ such that $\phi_* g_{ab} = g'_{ab}$. For ease of presentation, we will sometimes say that two manifolds $M$ and $M'$ are isometric when it is clear which metrics are associated with $M$ and $M'$. Two space-times $(M, g_{ab})$ and $(M', g'_{ab})$ are locally isometric if, for each point $p \in M$, there is an open neighborhood $O$ of $p$ and an open subset $O'$ of $M'$ such that $O$ and $O'$ are isometric and, correspondingly, with the roles of $(M, g_{ab})$ and $(M', g'_{ab})$ interchanged.

3. Observational Indistinguishability. We are now prepared to consider the notion of observationally indistinguishable space-times. Intuitively, one space-time is observationally indistinguishable from another if no observer in the first space-time has grounds for deciding which of the two she inhabits.3 Formally, we have the what follows.

**Definition 1.** Let $(M, g_{ab})$ and $(M', g'_{ab})$ be space-times. We say $(M, g_{ab})$ is observationally indistinguishable from $(M', g'_{ab})$ if, for every point $p \in M$, there is a point $p' \in M'$ such that $I^-(p)$ and $I^-(p')$ are isometric.

It should be clear from the definition that, given that a space-time $(M, g_{ab})$ is observationally indistinguishable from another space-time $(M', g'_{ab})$, no empirical data could (even in principle) allow an observer in $(M, g_{ab})$ to distinguish between the two models.

It has been shown that, except for certain pathological models, not only is every space-time observationally indistinguishable from some other, but the result holds even if we require the two space-times to have identical

---

3. There are various definitions of observational indistinguishability found in the literature. Here, we restrict our attention to one given by Malament (1977). For others, see Glymour (1972, 1977).
WHAT IS A PHYSICALLY REASONABLE SPACE-TIME?

local properties. In other words, cosmologists inherit a serious epistemic predicament even under the inductive assumption that “the normal physical laws we determine in our space-time vicinity are applicable at all other spacetime points” (Ellis 1975, 246). In order to present the result, consider the following definition.

**Definition 2.** A property $P$ on a space-time is **local** if, given any two locally isometric space-times $(M, g_{ab})$ and $(M', g'_{ab})$, $(M, g_{ab})$ has $P$ if and only if $(M', g'_{ab})$ has $P$. A property is **global** if it is not local.

Under this definition, it should be clear that, for example, the property of satisfying the standard energy conditions counts as local, while the property of being stably causal is classified as global.

Now, consider a special class of space-times $(M, g_{ab})$ that for some point $p$ in $M$, $I^-(p) = M$. Such space-times necessarily contain closed timelike curves and, in addition, have the property that all of space-time may be observed from one point. We will call these space-times **causally bizarre**.

Now we are ready to introduce the underdetermination result discussed above.

**Proposition 1.** Let $(M, g_{ab})$ be any space-time having any set of local properties $\mathcal{P}$. If $(M, g_{ab})$ is not causally bizarre, then there exists a space-time $(M', g'_{ab})$ such that (i) $(M', g'_{ab})$ has the set $\mathcal{P}$ of local properties, (ii) $(M, g_{ab})$ is observationally indistinguishable from $(M', g'_{ab})$, and (iii) $(M, g_{ab})$ and $(M', g'_{ab})$ are not isometric.

So, even if one fixes the local structure of space-time completely, there is a sense in which the global structure of every model is underdetermined. This poses quite a problem for the cosmologist. But, as we are about to see, the epistemic situation is even worse than the theorem above seems to suggest.

4. **Underdetermination of Global Properties.** In this section, we restrict our attention to four global properties of interest: inextendibility, isotropy, global hyperbolicity, and hole-freeness. Let us define each of them in turn.

We say a space-time $(M, g_{ab})$ is **inextendible** if it is not possible to properly embed it isometrically into another space-time. So, the property of inextendibility ensures that space-time is “as large as it can be.”

A space-time $(M, g_{ab})$ is (spatially) **isotropic** if there is a timelike vector field $\xi^a$ on $M$ such that, for all $p \in M$ and any two unit vectors $\sigma^a_1$ and $\sigma^a_2$ at $p$ that are orthogonal to $\xi^a$, there is an isometry $\varphi : M \to M$ that leaves $p$ and the field $\xi^a$ fixed but rotates $\sigma^a_1$ into $\sigma^a_2$. Intuitively, isotropic

4. See Wald (1984) for details concerning these properties.
5. The proof for this claim is given in Manchak (2009a).
space-times, sometimes called the “Friedmann models,” have no preferred spatial directions. A space-time \((M, g_{ab})\) is globally hyperbolic if there is a slice \(S\) in \(M\) such that \(D(S) = M\). Roughly, a globally hyperbolic space-time allows one to determine, from initial conditions on \(S\), the physical situation on all of \(M\). For this reason, global hyperbolicity is usually taken to be a necessary condition of Laplacian determinism.

A space-time \((M, g_{ab})\) is hole-free if, for any spacelike hypersurface \((M, g_{ab})\) there is no isometric embedding \(\theta: D(\Sigma) \rightarrow M'\) into another space-time \((M', g'_{ab})\) such that \(\theta(D(\Sigma)) \neq D(\theta(\Sigma))\). So, in a hole-free space-time, the Cauchy development of every spacelike surface is “as large as it can be.”

With these definitions in place, we are ready to strengthen the under-determination result from above. Let \(\Xi\) be the set of global properties consisting of inextendibility, isotropy, global hyperbolicity, and hole-freeness. We have what follows.

**Proposition 2.** Let \((M, g_{ab})\) be any space-time having any set of local properties \(\mathfrak{P}\). If \((M, g_{ab})\) is not causally bizarre, then there exists a space-time \((M', g'_{ab})\) such that (i) \((M', g'_{ab})\) has all of the properties in \(\mathfrak{P}\), (ii) \((M', g'_{ab})\) has none of the properties in \(\Xi\), (iii) \((M, g_{ab})\) is observationally indistinguishable from \((M', g'_{ab})\), and (iv) \((M, g_{ab})\) and \((M', g'_{ab})\) are not isometric.

*Proof.* Let \((M, g_{ab})\) be any space-time that is not causally bizarre having the set of local properties \(\mathfrak{P}\). Now construct \((M', g'_{ab})\) according to the method outlined in Manchak (2009a). Next, remove any point \(u\) in the \(M(1, \beta)\) portion of the manifold \(M'\). It is easily verified that the resulting space-time, call it \((M'', g''_{ab})\), is such that (i) \((M'', g''_{ab})\) has all of the properties in \(\mathfrak{P}\), (ii) \((M'', g''_{ab})\) has none of the properties in \(\Xi\), and (iii) \((M, g_{ab})\) is observationally indistinguishable from \((M', g'_{ab})\), and (iv) \((M, g_{ab})\) and \((M', g'_{ab})\) are not isometric. QED

We can understand the theorem to be saying that, not only is it impossible to fully establish the global structure of space-time, but we cannot even make partial determinations concerning a handful of space-time properties of interest. It seems that, although our universe may be inextendible, isotropic, globally hyperbolic, and hole-free, we can never know that it is. We mention in passing, however, that if it turns out that our universe fails to have the properties in \(\Xi\), it may be possible to determine that this is the case (see Malament 1977).

6. Note that, under this definition, spatial isotropy implies spatial homogeneity. See Ellis (2007, 1225).
5. Physical Reasonableness. Some may insist that the result just presented has little or no physical significance. One view is to hold that the space-time constructed in the proof cannot be physically reasonable because it was assembled via a seemingly artificial “cut-and-paste” process. But, such an objection must be formulated carefully because any space-time can be fabricated by such a construction (see Geroch 1971b, 78).

Some conditions for ruling out manufactured examples have been proposed. But, none seem to be entirely satisfactory. Inextendibility is not strong enough to forbid many cut-and-paste examples (see Earman 1995, 98). However, the property of local inextendibility, introduced by Hawking and Ellis (1973, 59), was later shown to be much too strong (see Beem 1980). And although the condition of hole-freeness seems to be adequate for many purposes, as we will see below, it carries with it some significant problems as well.

It must be remembered, too, that the theorem can have physical relevance, even if the particular model constructed in the proof does not. Geroch (1971b, 78) explains: “The space-times obtained by cutting and patching are not normally considered as serious models of our universe. However, the mere existence of a space-time having certain global features suggests that there are many models—some perhaps quite reasonable physically—with similar properties.” Some have held that any space-time that fails to have one or more of the properties in $\mathcal{S}$ is ipso facto physically unreasonable. In other words, the conditions of inextendibility, isotropy, global hyperbolicity, and hole-freeness have all been taken, at one time or another, to be satisfied by all reasonable models of our universe. But how does one justify such a position? We know that, given the theorem above, this justification cannot be due to any observational data we have collected or likewise any considerations of the local structure of space-time. In the remainder of this section, we will briefly examine the rationale in support of each of the properties under investigation. We hope to show that, in each case, the justification is dubious in certain respects.

Inextendibility. A space-time that fails to be inextendible is often thought to be physically unreasonable for metaphysical reasons (Earman 1995, 32). In particular, Leibniz’s principles of plenitude and sufficient reason seem to be at work. Geroch (1970, 262) asks, “Why, after all, would Nature stop building our universe . . . when She could just as well have carried on?” Others use similar reasoning (see Penrose 1969, 253; Clarke 1976, 17).

But, however compelling the metaphysics, it is sometimes problematic

7. See Norton (2008) for a related discussion of various ways a system can be regarded as “unphysical.”
to insist on inextendibility. For example, Clarke (1976, 20) has shown that not every well-behaved space-time admits a well-behaved inextendible extension. Should we cling to inextendibility at the expense of other desirable space-time properties? The answer is far from clear. Additionally, a space-time does not always have just one inextendible extension (Clarke 1993, 9). Thus, the principle of sufficient reason can actually be used to argue against the property of inextendibility. After all, why should one extension be preferred over another?

Isotropy. The Copernican principle is often used to support the claim that realistic models of the universe are (on sufficiently large scales) isotropic. Although precise formulations vary, this principle is generally taken to be the statement that we do not occupy a privileged position in space-time. Since observational data seem to indicate no preferred spatial direction from our vantage point, the Copernican principle implies an isotropic universe (see Wald 1984, 94). However, some have questioned the Friedmann models. Wainwright and Ellis (1996) have shown that “intermediate isotropisation” can occur in which space-time exhibits (highly approximate) isotropic behavior for arbitrarily long cosmic times, despite being extremely anisotropic at very early and very late epochs. In addition, we know that under the assumption of some formulations of cosmic inflation, we would not even expect our universe to be isotropic (Ellis 2007, 1227).

Elsewhere, it has been argued that although the Copernican principle seems to modestly deny us a special status, this “seeming modesty is belied by the immodest use to which the principle is put in justifying an inductive extrapolation” (Earman 1995, 129). Indeed, induction on such large scales would seem to be suspect, given that we are able to observe only a “negligibly small region” of the universe (see Wald 1984, 91).

Global Hyperbolicity. Motivated largely by considerations of causal determinism, some insist on the (strong) cosmic censorship hypothesis—the statement that all physically reasonable space-times are globally hyperbolic (see Joshi 1993; Earman 1995). To support such a position, Penrose (1979, 626) maintains that cosmological models that fail to be globally hyperbolic are unstable under certain types of perturbations. However, such a claim is difficult to express precisely (see Geroch 1971a). And, although some evidence does seem to indicate that instabilities are present in nonglobally hyperbolic space-times, still other evidence suggests otherwise (see Chandrasekhar and Hartle [1982] and Morris, Thorne, and Yurtsever [1988] for opposing perspectives).

A precise formulation of the cosmic censorship hypothesis due to Wald (1984, 304) eschews any reference to stability and, instead, simply forbids
known counterexamples. But, one must be careful with such an approach. As Earman (1995, 80) has emphasized, the term “physically unreasonable” should not be used as an “elastic label that can be stretched to include any ad hoc way of discrediting putative counterexamples.” Finally, even if a particular formulation can be agreed on, it seems that a proof of cosmic censorship is still a “long way” off (Penrose 1999, 245).

Hole-Freeness. The preservation of causal determinism also seems to be the motivating force behind the assumption that all physically reasonable space-times are hole-free. According to Clarke (1976, 17), hole-freeness is needed to ensure that predictions are “not falsified by the spontaneous appearance of uncaused singularities.” Additionally, Geroch (1977, 87) has suggested that, in order to uphold certain theorems concerning causal determinism, we may want to modify general relativity such that only hole-free space-times are permitted (see also Ellis and Schmidt 1977, 927). But, Earman (1995, 98) has argued that this imposition of hole-freeness amounts to little more than question begging. Indeed, it seems impermissible to justify the modification of one’s physical theories (so as to maintain determinism) merely because not doing so would put determinism in jeopardy.

But there are other, more serious, problems with hole-freeness. We know that hole-freeness can fail to mesh well with other properties of interest: not every hole-free space-time admits an inextendible hole-free extension (Clarke 1976, 20). And it has recently been shown that some inextendible, globally hyperbolic space-times are not hole-free (Manchak 2009b). In other words, hole-freeness is not necessarily a property of some maximal Cauchy developments. In fact, Krasinkov (2009) was able to show that even Minkowski space-time is not free of holes.

Under a more complicated alternative formulation—hole-freeness*—it does follow that every inextendible, globally hyperbolic space-time, including Minkowski space-time, is hole-free* (Manchak 2009b). But, given the concerns regarding inextendibility and global hyperbolicity mentioned above, it is still not clear whether this property of hole-freeness* must satisfied by all physically reasonable space-times. Indeed, there exist some (presumably physically reasonable) models of spherically symmetric, radiating stars, which are neither hole-free nor hole-free* (Steinmüller, King, and Lasota 1975).

6. Conclusion. It seems that one can certainly find principled reasons for believing that all physically reasonable cosmological models are inextendible, isotropic, globally hyperbolic, and hole-free. Indeed, metaphysical considerations, the Copernican principle, and a desire to preserve causal
determinism all lead one to certain conclusions concerning the global structure of space-time.

But as we have seen, these guiding principles, as well as the conclusions drawn from them, are far from uncontroversial. It is our position that the existence of this controversy, coupled with the underdetermination result above, requires a certain modesty with regard to the situation. One should simply be open to the possibility that our own universe is not best represented by an inextendible, isotropic, globally hyperbolic, hole-free model. Of course, this possibility demands that we apply great care when labeling as “physically unreasonable” space-times that fail to have some or all of these properties.

We close with some final thoughts concerning physical reasonableness and underdetermination. We have proceeded under the rather basic assumptions that space-time is temporally orientable, the manifold is Hausdorff, and so on. But, do we know that all physically reasonable space-times possess these properties? It seems that we do not—a construction similar to the one used above shows this.

So, it may be that our universe is not time orientable, or Hausdorff, and so on (see Earman 2002, 2008). But, the point is that we need not relax these standard assumptions of physicality in order to arrive at our findings. Indeed, we have attempted to strengthen our underdetermination results as much as possible by greatly limiting the class of admissible cosmological models and then showing that the epistemological predicament persists.

A few matters are still unresolved. Which global properties can be assumed while maintaining the underdetermination of space-time structure? In other words, for which sets of global properties \( \mathcal{G} \) will the following statement be true?

**Proposition 3.** Let \((M, g_{ab})\) be any space-time having any set of local properties \( \mathcal{B} \). If \((M, g_{ab})\) is not causally bizarre and has all of the properties in \( \mathcal{G} \), then there exists a space-time \((M', g'_{ab})\) such that (i) \((M', g'_{ab})\) has all of the properties in \( \mathcal{B} \) and \( \mathcal{G} \), (ii) \((M, g_{ab})\) is observationally indistinguishable from \((M', g'_{ab})\), and (iii) \((M, g_{ab})\) and \((M', g'_{ab})\) are not isometric.

One can straightforwardly show that if \( \mathcal{G} = \{ \text{inextendibility, stable causality} \} \), the proposition holds. What about other properties thought to be characteristic of physically reasonable space-time? In particular, if \( \mathcal{G} \) contains hole-freeness* or global hyperbolicity, does the underdetermination result fall apart? These questions remain open.

**REFERENCES**


