

What is a ‘Physically Reasonable’ Spacetime?*

John Byron Manchak

Abstract

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 a sense in which the global structure of spacetime cannot be fully established.¹ 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 upon 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 or not our universe is causally well-behaved in some respect. The first task of this paper 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 addition, these underdetermination results seem to have implications for what counts as a “physically reasonable” cosmological model.

*I wish to thank David Malament and John Norton for helpful comments.

¹See Manchak (2009a) for details.

Cosmologists often use certain global properties (for example, 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 paper is to carefully investigate this question.

2 Background structure

We start by reviewing the relevant background formalism of general relativity.² A relativistic *spacetime* is a pair of mathematical objects (M, g_{ab}) . M is a connected four-dimensional manifold (without boundary) that is smooth (infinitely differentiable). Here, g_{ab} is a smooth, non-degenerate, pseudo-Riemannian metric of Lorentz signature $(+, -, -, -)$ defined on M . Each point in the manifold represents an “event” in spacetime. For each point p in the manifold, the metric assigns a light cone structure in the tangent space M_p . Any tangent vector ξ^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* spacetime is one that has a continuous timelike vector field on M . A time orientable spacetime allows us to distinguish between the future and past lobes of the light cone. In what follows, we assume that spacetimes are time orientable.

For some interval $I \subseteq \mathbb{R}$, a smooth curve $\gamma : I \rightarrow M$ is *timelike* if the tangent vector ξ^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 γ 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 : $I^-(p) = \{q : q \ll p\}$, $I^+(p) = \{q : p \ll q\}$, $J^-(p) = \{q : q < p\}$, and $J^+(p) = \{q : p < q\}$. The set $I^-(p)$ will be used extensively throughout the paper and is called the *observational past* of p . This set represents those

²Details can be found in Hawking and Ellis (1973) and Wald (1984).

points which can possibly be observed from p .³

A point $p \in M$ is a *future endpoint* of a future-directed causal curve $\gamma : I \rightarrow M$ if, for every neighborhood O of p , there exists a point $t_0 \in I$ 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 $S \subset M$ is *achronal* if no two points in S can be connected by a timelike curve. The *edge* of a closed, achronal set $S \subset M$ 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 which does not intersect S . A set $S \subset M$ is a *slice* if it is closed, achronal, and without edge. A set $S \subset M$ is a *spacelike hypersurface* if S is an three-dimensional submanifold such that every curve in S is spacelike.

Two spacetimes (M, g_{ab}) and (M', g'_{ab}) are *isometric* if there is a diffeomorphism $\phi : M \rightarrow M'$ 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 spacetimes (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 spacetimes. Intuitively, a spacetime is observationally indistinguishable from another if no observer in the first spacetime has grounds for deciding which of the two she inhabits.⁴ Formally, we have the following.

³One uses the set $I^-(p)$ instead of $J^-(p)$ to represent the “observational past” of p for reasons of mathematical convenience.

⁴There 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).

Definition. Let (M, g_{ab}) and (M', g'_{ab}) be spacetimes. 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 spacetime (M, g_{ab}) is observationally indistinguishable from another spacetime (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 spacetime observationally indistinguishable from some other, but the result holds even if we require the two spacetimes to have identical 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 spacetime vicinity are applicable at all other spacetime points” (Ellis 1975, p. 246). In order to present the result, consider the following definition.

Definition. A property P on a spacetime is *local* if, given any two locally isometric spacetimes (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.⁵

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

Now we are ready to introduce the underdetermination result discussed above.⁶

Theorem. Let (M, g_{ab}) be any spacetime having any set of local properties \mathfrak{P} . If (M, g_{ab}) is not causally bizarre, then there exists a spacetime (M', g'_{ab}) such that (i) (M', g'_{ab}) has the set \mathfrak{P} of local properties, (ii) (M, g_{ab}) is ob-

⁵See Wald (1984) for details concerning these properties.

⁶The proof for this claim is given in Manchak (2009a).

servationally 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 spacetime 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 spacetime (M, g_{ab}) is *inextendible* if it is not possible to properly embed it isometrically into another spacetime. So, the property of inextendibility ensures that spacetime is “as large as it can be.”

A spacetime (M, g_{ab}) is (spatially) *isotropic* if there is a timelike vector field ξ^a on M such that, for all $p \in M$ and any two unit vectors σ_1^a and σ_2^a at p which are orthogonal to ξ^a , there is an isometry $\varphi : M \rightarrow M$ which leaves p and the field ξ^a fixed but rotates σ_1^a into σ_2^a . Intuitively, isotropic spacetimes, sometimes called the “Friedmann models”, have no preferred spatial directions.⁷

A spacetime (M, g_{ab}) is *globally hyperbolic* if there is a slice S in M such that $D(S) = M$. Roughly, a globally hyperbolic spacetime 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 spacetime (M, g_{ab}) is *hole-free* if, for any spacelike hypersurface Σ in M there is no isometric embedding $\theta : D(\Sigma) \rightarrow M'$ into another spacetime (M', g'_{ab}) such that $\theta(D(\Sigma)) \neq D(\theta(\Sigma))$. So, in a hole-free spacetime, 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 underdetermination result from above. Let \mathfrak{G} be the set of global properties consisting of inextendibility, isotropy, global hyperbolicity, and hole-freeness. We have

⁷Note that, under this definition, spatial isotropy implies spatial homogeneity. See Ellis (2007, p. 1225)

the following.

Theorem. Let (M, g_{ab}) be any spacetime having any set of local properties \mathfrak{P} . If (M, g_{ab}) is not causally bizarre, then there exists a spacetime (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 \mathfrak{S} , (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 spacetime which 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 spacetime, 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 \mathfrak{S} , and (iii) (M, g_{ab}) is observationally indistinguishable from (M'', g''_{ab}) , and (iv) (M, g_{ab}) and (M'', g''_{ab}) are not isometric. \square

We can understand the theorem to be saying that, not only is it impossible to fully establish the global structure of spacetime, but we cannot even make partial determinations concerning a handful of spacetime 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 \mathfrak{S} , it may be possible to determine that this is the case (see Malament 1977).

5 Physical Reasonableness

Some may insist that the result just presented has little or no physical significance.⁸ One view is to hold that the spacetime 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* spacetime can be fabricated by such a construction (see Geroch 1971a, p. 78).

⁸See Norton (manuscript) for a related discussion of various ways a system can be regarded “unphysical”.

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, p. 98). On the other hand, the property of local inextendibility, introduced by Hawking and Ellis (1973, p. 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 (1971a, p. 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 spacetime which fails to have one or more of the properties in \mathfrak{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 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 spacetime. 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 spacetime which fails to be inextendible is often thought to be physically unreasonable for metaphysical reasons (Earman 1995, p. 32). In particular, Leibniz’s principles of plenitude and sufficient reason seem to be at work. Geroch (1970, p. 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, p. 253; Clarke 1976, p. 17).

But, however compelling the metaphysics, it is sometimes problematic to insist on inextendibility. For example, Clarke (1976, p. 20) has shown that not every well-behaved spacetime admits a well-behaved inextendible extension. Should we cling to inextendibility at the expense of other desirable

spacetime properties? The answer is far from clear. Additionally, a spacetime does not always have just one inextendible extension (Clarke 1993, p. 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 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 spacetime. Since observational data seem to indicate no preferred spatial direction from our vantage point, the Copernican principle implies an isotropic universe (see Wald 1984, p. 94). However, some have questioned the Friedmann models. Wainwright and Ellis (1996) have shown that “intermediate isotropisation” can occur in which spacetime 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, p. 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, p. 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, p. 91).

Global Hyperbolicity. Motivated largely by considerations of causal determinism, some insist on the (strong) cosmic censorship hypothesis – the statement that all physically reasonable spacetimes are globally hyperbolic (see Joshi 1993; Earman 1995). To support such a position, Penrose (1979, p. 626) maintains that cosmological models which fail to be globally hyperbolic are unstable under certain types of perturbations. However, such a claim is difficult to express precisely (see Geroch 1971b). And, although some evidence does seem to indicate that instabilities are present in non-globally hyperbolic spacetimes, 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, p. 304) eschews any reference to stability and, instead, simply forbids known counterexamples. But, one must be careful with such an approach. As

Earman (1995, p. 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 upon, it seems that a proof of cosmic censorship is still a “long way” off (Penrose 1999, p. 245).

Hole-Freeness. The preservation of causal determinism also seems to be the motivating force behind the assumption that all physically reasonable spacetimes are hole-free. According to Clarke (1976, p. 17), hole-freeness is needed to ensure that predictions are “not falsified by the spontaneous appearance of uncaused singularities.” Additionally, Geroch (1977, p. 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 spacetimes are permitted (see also Ellis and Schmidt 1977, p. 927). But, Earman (1995, p. 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.

Other problems are also associated with hole-freeness. It has recently been shown that, contrary to claims made in the literature, some inextendible, globally hyperbolic spacetimes are not hole-free (Manchak 2009b). In other words, hole-freeness is not necessarily a property of some maximal Cauchy developments. Also, there exist some (presumably physically reasonable) models of spherically symmetric, radiating stars which are not hole-free (Steinmüller, King, and Lasota 1975). Finally, we know that hole-freeness can fail to mesh well with other properties of interest: not every hole-free spacetime admits an inextendible hole-free extension (see Clarke 1976, p. 20).

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 spacetime.

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” spacetimes which fail to have some or all of these properties.

References

- [1] Beem, J. (1980), “Minkowski Space-Time is Locally Extendible”, *Communications in Mathematical Physics*, 72: 273-275.
- [2] Chandrasekhar, S., and J. B. Hartle (1982), “On Crossing the Cauchy Horizon of a Reissner-Nordström Black-Hole”, *Proceedings of the Royal Society (London) A*, 348: 301-315.
- [3] Clarke, C. J. S. (1976), “Spacetime Singularities”, *Communications in Mathematical Physics*, 49: 17-23.
- [4] Clarke, C. J. S. (1993), *The Analysis of Space-Time Singularities*. Cambridge: Cambridge University Press.
- [5] Earman, J. (1995), *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. New York: Oxford University Press.
- [6] Ellis, G. F. R. (1975), “Cosmology and Verifiability”, *Quarterly Journal of the Royal Astronomical Society*, 16: 245-264.
- [7] Ellis, G. F. R. (2007), “Issues in the Philosophy of Cosmology”, in J. Butterfield and J. Earman (eds.), *Handbook of the Philosophy of Physics*. Oxford: Elsevier, 1183-1286.
- [8] Ellis, G. F. R., and Schmidt, B. G. (1977), “Singular Spacetimes”, *General Relativity and Gravitation*, 8: 915-953.
- [9] Geroch, R. (1970), “Singularities”, in M. Carmeli, S. I. Fickler, and L. Witten (eds.), *Relativity*. New York: Plenum Press, 259-291.

- [10] Geroch, R. (1971a), “Spacetime Structure from a Global Viewpoint”, in B. K. Sachs (ed.), *General Relativity and Cosmology*. New York: Academic Press, 71-103.
- [11] Geroch, R. (1971b), “General Relativity in the Large”, *General Relativity and Gravitation*, 2: 61-74.
- [12] Geroch, R. (1977), “Prediction in General Relativity”, in J. Earman, C. Glymour, and J. Stachel (eds.), *Foundations of Space-Time Theories*. Minnesota Studies in the Philosophy of Science, vol. 8. Minneapolis: University of Minnesota Press, 81-93.
- [13] Geroch, R., and G. Horowitz (1979), “Global Structure of Spacetimes”, in S. Hawking and W. Israel (eds.), *General Relativity: An Einstein Centenary Survey*. Cambridge: Cambridge University Press, 212-293.
- [14] Glymour, C. (1972), “Topology, Cosmology, and Convention,” *Synthese*, 24: 195-218.
- [15] Glymour, C. (1977), “Indistinguishable Space-Times and the Fundamental Group”, in J. Earman, C. Glymour, and J. Stachel (eds.), *Foundations of Space-Time Theories*. Minnesota Studies in the Philosophy of Science, vol. 8. Minneapolis: University of Minnesota Press, 50-60.
- [16] Hawking, S. W., and G. F. R. Ellis (1973), *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press.
- [17] Joshi, P. (1993), *Global Aspects in Gravitation and Cosmology*. Oxford: Oxford University Press.
- [18] Malament, D. (1977), “Observationally Indistinguishable Space-Times”, in J. Earman, C. Glymour, and J. Stachel (eds.), *Foundations of Space-Time Theories*. Minnesota Studies in the Philosophy of Science, vol. 8. Minneapolis: University of Minnesota Press, 61-80.
- [19] Manchak, J. B. (2009a), “Can We Know the Global Structure of Spacetime?”, *Studies in History and Philosophy of Modern Physics*, 40: 53-56.
- [20] Manchak, J. B. (2009b), “Is Spacetime Hole-Free?”, *General Relativity and Gravitation*, forthcoming.

- [21] Morris, M. S., K. S. Thorne, and U. Yurtsever (1988), “Wormholes, Time Machines, and the Weak Energy Condition”, *Physical Review Letters*, 61: 1446-1449.
- [22] Norton, J. (manuscript), “The Dome: An Unexpectedly Simple Failure of Determinism.”
- [23] Penrose, R. (1969) “Gravitational Collapse: The Role of General Relativity,” *Revisita del Nuovo Cimento*, Serie I, 1: 252-276.
- [24] Penrose, R. (1979), “Singularities and Time-Asymmetry”, in Stephen Hawking and Werner Israel (eds.), *General Relativity: An Einstein Centenary Survey*. Cambridge: Cambridge University Press, 581-638.
- [25] Penrose, R. (1999), “The Question of Cosmic Censorship”, *Journal of Astrophysics and Astronomy*, 20: 233-248.
- [26] Steinmüller, B., A. R. King, and J. P. Lasota (1975), “Radiating Bodies and Naked Singularities”, *Physics Letters A*, 51: 191-192.
- [27] Wainwright, J., and G. F. R. Ellis (1996), *The Dynamical Systems Approach to Cosmology*. Cambridge: Cambridge University Press.
- [28] Wald, R. (1984), *General Relativity*. Chicago: University of Chicago Press.