# Determinism and General Relativity

Chris Smeenk and Christian Wüthrich*†

We investigate the fate of determinism in general relativity (GR), comparing the philosopher's account with the physicist's well-posed initial value formulations. The fate of determinism is interwoven with the question of what it is for a spacetime to be 'physically reasonable'. A central concern is the status of global hyperbolicity, a putatively necessary condition for determinism in GR. While global hyperbolicity may fail to be true of all physically reasonable models, we analyze whether global hyperbolicity should be (i) imposed by fiat; (ii) established from weaker assumptions, as in cosmic censorship theorems; or (iii) justified by beyond-GR physics.

**1. Introduction.** Two foundational questions one might ask about any physical theory bring out particularly subtle and interesting features of general relativity (GR). First, is GR a deterministic theory? Second, do all mathematical models of the theory represent physically possible spacetimes? There is a tight connection in GR between these two questions, that is, between an assessment of what spacetimes are physically possible or reasonable and of whether determinism holds. It is this connection that we explore in this essay.

Determinism holds if specifying the state of a system uniquely fixes its dynamical evolution. Spacetimes with exotic causal structure raise a distinctive set of questions regarding the status of determinism in GR, differing from those raised by the (in)famous hole argument. Below we bypass the hole argument by assuming that the existence of a unique solution 'up to diffeomorphism invariance' is sufficient for determinism in GR. We consider cases

Received June 2020; revised September 2020.

*To contact the authors, please write to: Chris Smeenk, Rotman Institute of Philosophy and Department of Philosophy, University of Western Ontario; e-mail: csmeenk2@uwo.ca. Christian Wüthrich, Department of Philosophy, University of Geneva; e-mail: christian.wuthrich@unige.ch.

†We thank Gordon Belot, Juliusz Doboszewski, J. B. Manchak, the audience at PSA 2010 in Montreal, the New Directions group at Western University, and two anonymous referees for their valuable feedback and exchanges. We acknowledge partial support from the John Templeton Foundation (grants 61048 and 61387). Both authors contributed equally to this article.

Philosophy of Science, 88 (October 2021) pp. 638–664. 0031-8248/2021/8804-0004$10.00
Copyright 2021 by the Philosophy of Science Association. All rights reserved.



https://doi.org/10.1086/713904 Published online by Cambridge University Press

in which uniqueness fails regardless of the interplay between metaphysics and diffeomorphism invariance.[1] In the cases we are interested in, fully specifying the state of the universe within a spacetime region fails to fix the maximal extension of the region. Determinism fails in the straightforward sense that everything within the region is compatible with multiple extensions beyond it. Even though the state may in some sense locally determine a unique solution, it fails to fix a unique global, maximal solution.

Our analysis runs parallel to recent discussions of determinism in Newtonian particle mechanics. Philosophers have considered unusual cases illustrating several distinct ways in which determinism can fail in Newtonian mechanics. For example, if we allow force functions that are not continuously differentiable, Newtonian dynamics admits multiple solutions for the same initial data.[2] But are these 'physically reasonable' forces, falling within the domain of applicability of Newtonian theory? Similarly, many solutions of Einstein's field equation (EFE) have unusual mathematical features that seem unreasonable in some sense and apparently do not describe the actual universe or a region of it. How do we decide to treat some solutions as beyond the pale? As with Malament's (2008) take on whether Newtonian mechanics is deterministic, our interest is in clarifying the basis for such judgments rather than rendering a verdict on determinism.[3]

Making the world safe for determinism in GR requires substantive claims regarding the global structure of space and time along with constraints on the permissible types of matter. The conditions needed to reject some solutions as 'unphysical' are quite different in character from the local dynamical laws. Although the content of these conditions is clear, their status is not. We will discuss three different approaches below that reflect different reasons for placing tighter constraints on physical possibility than those implied by EFE alone. The first approach offers a modern transcendental argument: although GR dispenses with much 'background structure', there are still some global conditions on the spacetime manifold that are prerequisites for sensible physics. In this

1. None of this is to deny the interest of the questions raised by the hole argument (Earman and Norton 1987; see also Stachel [2014] for a recent review). For reasons that become clear below, we are sympathetic to the position, defended recently by Weatherall (2018), that we should always take mathematical models to be only individuated 'up to isomorphism' for the relevant mathematical objects.

2. Several recent discussions have focused on a simple case: a particle moving along Norton's dome (Norton 2003), a constraint surface that does not have a well-defined second derivative at its apex. A similar pathology arises for the following force function: $f(r) = m\sqrt{\ell - r}$ for $r \leq \ell$, and $= 0$ for $r > \ell$. Because $f(r)$ is not differentiable at $r = \ell$, Newtonian dynamics fails to admit a well-posed initial value formulation (IVF). See Fletcher (2012) for a thorough review, who notes in passing that the provenance of these problems goes back to Poisson in the early nineteenth century.

3. Throughout this article, we draw on the seminal survey of related topics in Earman (1995).

https://doi.org/10.1086/713904 Published online by Cambridge University Press

spirit, conditions closely tied to determinism may turn out to be necessary in order for GR to get any traction at all and then be parlayed into an argument to the conclusion that GR is deterministic. The second approach shifts the goal posts: rather than attempting to prove determinism in an entirely general sense, cosmic censorship theorems aim to show that a suitably restricted subset of solutions satisfies a stability property. Since stability under small perturbations is arguably a property of 'physically reasonable' spacetimes, this approach offers a second avenue from physical reasonableness to determinism. Finally, a third approach emphasizes that our assessment of 'physical possibility' must take into account the limits of applicability of classical GR and the implications of other theories. We briefly discuss the possibility that a successor theory may rule out the failures of determinism allowed in classical GR and the related possibility that global conditions may serve as preconditions for the possibility of formulating quantum field theory (QFT).

The present analysis bears on an important debate in the foundations of physics. One side in this debate considers determinism to be a basic *principle* crucial to physicists. Advocates of this stance often take it to be part of clarifying the content of a theory to see what is required to satisfy determinism. The other side insists that there cannot be such an a priori principle and typically treats the various ways of insuring determinism as a possibly illegitimate exercise. Our purpose here is not to argue for a position in this debate but to clarify the distinctive issues that arise in GR. We agree with Earman (2004) that those who expect a simple answer from modern physics to the question of whether the world is deterministic will be frustrated, as we have also argued elsewhere (Earman, Smeenk, and Wüthrich 2009; Smeenk and Wüthrich 2011; Wüthrich 2011). The foundations of modern physics are simply too rich for such a simple-minded expectation to be borne out.

We will start in section 2 by fixing the terminology and by discussing the implications of moving from the parabolic partial differential equations (PDEs) of prerelativistic physics to the hyperbolic ones in relativity. In section 3, we develop the conception of determinism in the context of GR. In section 4, we consider whether global hyperbolicity is necessary or sufficient as a condition for determinism in GR. In sections 5, 6, and 7, respectively, we will discuss three approaches to conceiving of the status of global hyperbolicity as a condition (i) to be stipulated a priori, (ii) to be proved from weaker assumptions within GR, and (iii) to be confirmed—or infirmed—from physics beyond GR. Brief conclusions follow in section 8.

**2. Hyperbolic PDEs and Relativity.** Let us start from the idea of determinism as requiring that suitably specifying the physical state of a system in some region of spacetime $S$ uniquely fixes, through the dynamics, the state throughout a larger region $S'$ ($S \subset S'$)—in its strongest form, throughout the entire spacetime. Physical theories typically describe dynamical

https://doi.org/10.1086/713904 Published online by Cambridge University Press

evolution using PDEs. In this context, the status of determinism translates into the question of whether the system of PDEs specifying dynamical evolution, supplemented with some further input (information regarding the state of the system in $S$), yields a unique solution throughout $S'$. What form the input must take to lead to a uniquely solvable problem, and the qualitative features of the solutions, differs among the three classes of PDEs—hyperbolic, parabolic, and elliptic. Over the course of the twentieth-century physics has shifted to using hyperbolic PDEs to represent fundamental interactions, with profound consequences for the status of determinism, as the following considerations illustrate.[4]

One challenge to determinism in Newtonian physics stems from the fact that the relevant equations do not place an upper limit on signal propagation speed (see Earman 1986, 2004). (For Galilean invariant laws, one can always choose an inertial reference frame with respect to which a given particle has an arbitrarily high velocity.) This raises the possibility of 'space invaders': particles that 'fly in from infinity' and disrupt an otherwise well-behaved system. Xia (1992) showed that this possibility can be realized in a system of 5 point particles interacting through Newtonian gravity. The particles can be arranged such that their trajectories all approach a given time slice asymptotically: the particles either show up 'out of nowhere' or, considering the time-reversed trajectories, 'fly off to infinity' within a finite time. In light of such behavior, the region of spacetime $S$ appropriate for setting up a uniquely solvable problem has to include the boundaries of the system over time—specifying, for example, that the number of particles remains fixed. Parabolic PDEs, such as the heat equation, similarly require specifying boundary values to find a unique solution. For so-called initial-boundary value problems, the region $S$ where input data need to be supplied consists of a region $\Sigma$ at a given time along with its boundary over time (see fig. 1 for an example).

Hyperbolic PDEs have strikingly different qualitative behavior that is a boon to determinism. The familiar wave equation, $\partial_x^2 u = \partial_t^2 u$, exhibits features that hold for this general class of PDEs. Solutions are 'localized' in that waves produced by a disturbance propagate at finite speed, along so-called characteristic curves, in contrast to the infinite-speed propagation of effects allowed by both parabolic and elliptic PDEs. The appropriate surface for specifying the input data is a spatially extended region $S$ at a given time. As a result of the finite propagation speed, there are nontrivial domains of

4. Obviously parabolic and elliptic PDEs are still used in a wide variety of modeling contexts. Here we follow Geroch (1996, 20), who remarks that "a case could be made that, at least on a fundamental level, all the 'partial differential equations of physics' are hyperbolic—that, e.g., elliptic and parabolic systems arise in all cases as mere approximations of hyperbolic systems."

https://doi.org/10.1086/713904 Published online by Cambridge University Press

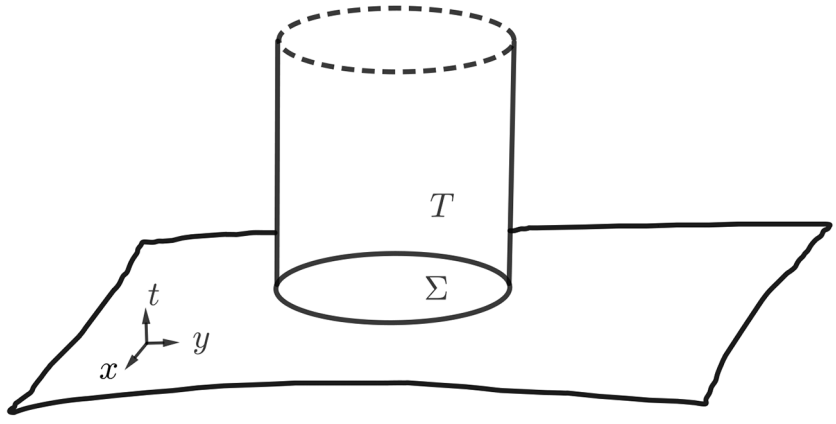



Figure 1. In this example of an initial-boundary value problem, the region of input data consists of the finite region at a given time Σ and its cylindrical boundary over time, *T*.

dependence—roughly speaking, the region of spacetime that can be connected to *S* by signals going at or below the wave propagation speed. For any point falling within the domain of dependence, all waves propagating to or from the point must 'register' on *S*. The state of the system throughout this larger region is fixed by the wave equation along with the initial data on *S*. There is no need to supplement the data on *S* with conditions regarding how the solution behaves 'on the edge' of *S* or at other times. Maxwell's equations share these features of the wave equation (albeit with additional complications due to gauge freedom). Taking the causal structure of these equations to be universal rather than characteristic of electromagnetism leads to special relativity. Minkowski spacetime provides a natural setting for formulating hyperbolic PDEs, as the causal structure of the spacetime reflects the existence of a maximum propagation speed.

We need to introduce some technical machinery to make this more precise. A *spacetime* is an ordered pair $\langle \mathcal{M}, g_{ab} \rangle$, where $\mathcal{M}$ is a connected four-dimensional differentiable manifold without boundary, and $g_{ab}$ is a Lorentz-signature metric defined everywhere on $\mathcal{M}$ that satisfies EFE—there is a stress-energy tensor $T_{ab}$ and value of the cosmological constant $\Lambda$ such that $R_{ab} - (1/2)Rg_{ab} - \Lambda g_{ab} = 8\pi T_{ab}$. This last clause only has some bite if there are further restrictions on $T_{ab}$. Some results regarding causal structure follow without placing any constraints on $T_{ab}$. Most of the discussion below will focus on either exact solutions, with a specific choice of $T_{ab}$ representing particular type(s) of matter fields, or 'generic' solutions for which $T_{ab}$ satisfies constraints such as the energy conditions (see also Malament 2012, 159–61).

https://doi.org/10.1086/713904 Published online by Cambridge University Press

The metric $g_{ab}$ fixes a light cone structure in the tangent space $\mathcal{M}_p$ at each point $p \in \mathcal{M}$. A tangent vector $\xi^a \in \mathcal{M}_p$ is classified as time-like, null, or space-like, according to whether $g_{ab}\xi^a\xi^b > 0$, $= 0$, or $< 0$ (respectively). Geometrically, the null vectors 'form the light cone' in the tangent space with the time-like vectors lying inside and the space-like vectors lying outside the cone. The classification of tangent vectors extends to curves, and for a time-orientable spacetime, which admits a globally consistent designation of the 'future' lobe of the null cone, one can further define future-directed time-like (causal) curves as those whose tangent vectors fall within (on or within) the future lobe at each point.

The light cone structure and classification of curves leads directly to more general characterizations of the causal structure, including domains of dependence. The *chronological past* $I^-(p)$ is defined as the set of all points in $\mathcal{M}$ that can be reached from $p$ by past-directed time-like curves (analogously for the chronological future $I^+(p)$). For an *achronal surface* $V$, there are no points $p, q \in V$ such that $p \in I^-(q)$. The *edge* of an achronal surface $V$ is the set of points $p$ such that every open neighborhood $O$ of $p$ includes points in $I^+(p)$ and $I^-(p)$ that can be connected by a time-like curve that does not cross $V$. A *global time slice* of a spacetime $\langle \mathcal{M}, g_{ab} \rangle$ is a closed, achronal hypersurface $\Sigma \subset \mathcal{M}$ with no edges. The *future domain of dependence* $D^+(\Sigma)$ for a global time slice $\Sigma$ includes the points in $\mathcal{M}$ such that every past inextendible causal curve through $p$ intersects $\Sigma$ (analogously for the past domain of dependence $D^-(\Sigma)$).[5] A spacetime $\langle \mathcal{M}, g_{ab} \rangle$ is *globally hyperbolic* just in case there is a slice $\Sigma$ such that the domain of dependence $D(\Sigma) := D^+(\Sigma) \cup D^-(\Sigma)$ is the entire spacetime (i.e., $D(\Sigma) = \mathcal{M}$).

With these definitions in place, the contrast between Newtonian physics and special relativity can be stated simply: input data on a time slice $\Sigma$ in Newtonian spacetime gets us nowhere, in the sense that $D^+(\Sigma) = D^-(\Sigma) = \Sigma$, whereas Minkowski spacetime allows for the possibility of nontrivial domains of dependence. Whether this possibility is realized depends on whether the matter fields being studied respect the causal structure. In particular, matter fields that satisfy the *dominant energy condition* do not propagate outside the light cone.[6] Proofs that matter fields admit a well-posed IVF in Minkowski spacetime depend on what are called 'energy estimates'. Consider slices $\Sigma_1$ and $\Sigma_2$, at a later time, and the regions $S \subset \Sigma_1$ and $T = D^+(S) \cap \Sigma_2$. For matter fields satisfying the dominant energy condition, the field equations imply

5. A past inextendible causal curve $\gamma(s) : I \to \mathcal{M}$ does not have a past endpoint, namely, a point $p \in \mathcal{M}$ such that for every neighborhood $O$ of $p$, there is a value $s_0 \in I$ such that $\gamma(s) \in O$ for all values $s < s_0$.

6. A matter field with stress-energy tensor $T_{ab}$ satisfies the dominant energy condition if $T_{ab}\xi^a\xi^b \geq 0$ and $T^a_b\xi^b$ is time-like or null, given an arbitrary time-like vector field $\xi^a$ at any point in $\mathcal{M}$ (see, e.g., Malament 2012, 144–45).

https://doi.org/10.1086/713904 Published online by Cambridge University Press

bounds on the energy associated with initial data on $S$ and $T$.[7] To establish uniqueness of a solution to a linear wave equation for specified initial data, schematically, suppose for reductio that there are two solutions with the same initial data on $S$. Consider a third solution defined as their difference. The energy of this third solution on $S$ is zero, by hypothesis, but it follows from the energy bounds that this quantity cannot differ from zero on $T$ for an arbitrary later slice $\Sigma_2$. Hence, the two solutions must be identical within $D^+(S)$. Energy estimates also play an essential role in establishing the other two properties required for a well-posed IVF, namely, that solutions exist and vary continuously with the initial data.[8] For linear wave equations, with suitable initial data specified on a time slice $\Sigma$ (a.k.a. a Cauchy surface), $D(\Sigma) = \mathcal{M}$.[9] Hence, the transition to Minkowski spacetime supports the claim that fundamental physics satisfies determinism.

The consequences of the further step from special relativity to GR are more challenging to assess. To represent the gravitational field via spacetime geometry, the light cone structure of Minkowski spacetime has to be demoted from global to local status. Modern presentations of GR sometimes generalize this claim: the mathematical structures typically treated as global, background features of spacetime prior to GR should be 'stripped away', leaving only a bare differentiable manifold. In GR the causal structure is itself dynamical rather than fixed, leaving us unable to appeal to various background structures used in the analysis of other PDEs. A further source of complexity is the nonlinearity of EFE. Generally speaking, for nonlinear field equations 'good' initial data may lead to singularities, where the derivatives 'blow up' and existence and uniqueness results fail within a finite time—as illustrated by the formation of shock fronts in fluid mechanics (Christodoulou 2007). This possibility poses a general obstacle to understanding the 'global' behavior of solutions even if the dynamics admits a well-posed IVF 'locally' (close to the initial data surface, in a precise sense). In addition, although we noted above that we will set aside issues related to the hole argument, the

7. The energy over a space-like region $S$ is obtained by integrating $T_{ab}\xi^a\xi^b$ over $S$, where $\xi^a$ is a time-like vector field normal to $S$. Often the 'energy' used in PDE techniques does not, however, correspond directly to a physically meaningful quantity.

8. See sec. 7.4 of Hawking and Ellis (1973) for a proof that quasi-linear (linear in the highest derivatives), second-order hyperbolic systems admit a well-posed IVF. The dominant energy condition is not necessary: in some cases when $T_{ab}$ fails to satisfy this condition, it still satisfies energy inequalities sufficient to prove that there is a well-posed IVF.

9. The proofs for linear wave equations have been generalized to some nonlinear systems, but it is much more challenging to prove global existence theorems for these cases, and results are limited. Recent work has also focused on proving theorems given weaker assumptions about the regularity of the initial data and solution spaces (see Rendall 2008; Ringström 2009).

https://doi.org/10.1086/713904 Published online by Cambridge University Press

diffeomorphism invariance of GR leads to complications in providing an IVF. One complication resembles that faced by theories with gauge freedom, such as Maxwell's equations: the initial data cannot be freely specified but must satisfy a set of constraint equations, elliptic PDEs. All existing proofs of the existence and uniqueness of solutions for EFE use additional structures or coordinate conditions, rather than employing a framework that maintains manifest diffeomorphism invariance throughout. Techniques called hyperbolic reduction, for example, can be employed to derive a system of hyperbolic PDEs from EFE. As with Maxwell's equations, these have to be accompanied by elliptic constraint equations, and one has to furthermore ensure that the constraints propagate (i.e., that they continue to hold under the 'time evolution' equations extracted from EFE). Yet other formulations reduce the EFE to a system of parabolic equations or lead to constraint equations that are ordinary differential equations (as in the characteristic initial value problem).[10]

These aspects of GR combine to make the question of determinism so intriguing. There is much more work in mathematical physics needed to elucidate the structure of second-order nonlinear PDEs such as EFE. An understanding of the status of determinism in GR comparable to what we have achieved in other physical theories depends on the outcomes of this work. But even as we await resolution of these issues, we might aim as a modest first step to distinguish between failures of determinism that arise due to what looks like a mistake in mathematics, namely, stripping away structures that turn out to be essential to formulating physics, and those that reflect underlying features of the dynamics, such as the formation of 'naked singularities'.

**3. Defining Determinism.** With the terminology and the background in place, let us consider various proposals for defining determinism in the context of GR. Budic and Sachs (1976, 22) define a spacetime to be deterministic just in case $\forall\, p \in \mathcal{M}$, every past-endless causal curve that intersects $I^+(p)$ also intersects $I^-(p)$.[11] As Budic and Sachs themselves admit, this is a very restrictive definition; they even prove that no asymptotically flat spacetime is deterministic in their sense. Among many other important models, it classifies Minkowski spacetime as nondeterministic. However, they defend their characterization as adequate because it captures the intuitive sense that in deterministic spacetimes "each event can, in principle, predict from its own past everything [i.e., its chronological future], not merely its own future" (23).

10. For further discussion of the different formulations, see, e.g., Friedrich and Rendall (1999) and Ringström (2009).

11. A causal curve is *past-endless* just in case there is no event $p \in M$ such that going in the past direction, for every neighborhood $U(p)$ of $p$, the curve enters and remains in $U$.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

We do not share this intuition. In fact, we find it curious that anyone would think that the past of a single event would suffice to fix its complete (chronological) future in a dynamical system where effects propagate at finite speed, that is, in a system described by hyperbolic PDEs. Our position here is supported by results that have become known under the label of 'observational indistinguishability', to be discussed in section 5, which establish that the past of an event does not, quite generally and under relatively mild assumptions, fix the rest of the spacetime. Furthermore, the proposal blurs the crucial distinction between predictability and determinism.[12] On our view, the fate of determinism does not depend on whether the input data needed to specify a uniquely solvable problem are accessible to an embedded observer.

The conception of determinism we are seeking instead follows the standard Laplacian notion, at least in spirit.[13] The Laplacian notion encodes a sense of metaphysical determination and is deprived of epistemic undertones. It captures the idea that, given the state of the physical system at stake at a time, the laws of nature as encoded in the theory describing the system uniquely determine the state of the system at all other times. More precisely, for a globally deterministic theory, for appropriate initial data specified within an appropriate spacetime region $S$ there is a unique solution of the relevant equations (including EFE and equations governing any coupled matter fields) through the entire spacetime ($D(S) = \mathcal{M}$). A more rigorous definition would require filling in the two instances of 'appropriate' in our characterization. For vacuum solutions, the initial data for gravitational degrees of freedom typically include the induced three-dimensional metric $h_{ab}$ and the extrinsic curvature $k_{ab}$ on a three-dimensional hypersurface.[14] More generally, the form initial data must take in order to be 'appropriate' will depend on the theory of matter to be used.[15] For hyperbolic PDEs in special relativity,

12. This is not to deny the interest of questions regarding predictability. See, e.g., Geroch (1977).

13. Butterfield (1989) offers a relativistic conception of determinism, further generalized by Doboszewski (2019, 11). Although we heartily endorse Doboszewski's (2019) recognition that there are a variety of further conditions that might be imposed in defining determinism, we hold that there is still a core conception underlying these various modifications.

14. Provided that they satisfy the Gauss-Codacci constraints. Although the most familiar IVF for EFE uses initial data of this type, there are other possibilities. (We will mention one of these below; see Sarbach and Tiglio [2012] for discussion and further references.)

15. For typical field equations, the initial data include the field values and their first derivatives (with respect to the 'time' coordinate orthogonal to the surfaces $S$). There are further choices to be made regarding what function space the initial data belong to, on the basis of the degree of regularity required (for both metric and matter degrees of freedom). Oddly, proofs of uniqueness require a higher degree of differentiability than the existence results.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

initial data can be assigned to space-like hypersurfaces of constant time, so a natural candidate for the 'appropriate' region is the relativistically next best thing to a moment of time—that is, a global time slice Σ. (But all time slices are not created equal. We further need to rule out surfaces such as an achronal surface Σ for which $\exists\, p : \Sigma \subset I^{-}(p)$ in Minkowski spacetime; for such a surface, $D(\Sigma)$ is not the entire manifold, but this reflects a poor choice of Σ.)

A theorem due to Choquet-Bruhat and Geroch (1969) establishes that appropriate initial data $(\Sigma, h_{ab}, k_{ab})$, specified on a global time slice Σ, determine a unique (up to isometry), maximal globally hyperbolic development (MGHD), a solution to vacuum EFE. If matter fields are included, the initial data will consist of two types of information, specifying the state of the spacetime geometry and the matter fields, respectively, and the dynamical equations at play will then include EFE as well as the dynamical equations of the matter degrees of freedom. It is generally expected that matter fields that have a well-posed IVF in special relativity also have one in this context.[16] The restriction of the result to the globally hyperbolic development is as essential as it is unsurprising: regions that are not globally hyperbolic fall outside of $D(\Sigma)$ by definition. The solution determined by the initial data is the MGHD of the data, but it is not the maximal development simpliciter (a point we return to below).

We should pause to note a subtle contrast between the treatment of an IVF for GR and for other PDEs. The background spacetime structure is not given ab initio, and we currently lack mathematical tools that would allow for a fully diffeomorphism-invariant treatment. What we have instead are a variety of methods for reducing the full EFE to a system of constraint and evolution equations. Choquet-Bruhat discovered that existence and uniqueness results from hyperbolic PDEs could be extended to EFE, which then take the form of a nonlinear wave equation, given initial data $(\Sigma, h_{ab}, k_{ab})$ satisfying constraint equations. But this data set $(\Sigma, h_{ab}, k_{ab})$ has to be regarded as defined on an 'abstract' space—we cannot, without begging the question, assume that Σ fits within a solution. The existence proof establishes that there is an embedding map $\phi : \Sigma \to \mathcal{M}$ such that $h_{ab}, k_{ab}$ appropriately match the spacetime metric $g_{ab}$, where $\langle \mathcal{M}, g_{ab} \rangle$ is a solution to EFE. Turning to uniqueness, we can further ask whether the same initial data set can be embedded in two distinct solutions. The proof of uniqueness establishes, roughly put, that any such pair of solutions must be extensions of one underlying solution

16. The proof mentioned in n. 8 carries over to the coupled Einstein-matter field equations for many types of matter fields, granting some plausible assumptions regarding the coupling between the matter fields and gravitational degrees of freedom. But there are numerous open questions, related, e.g., to the treatment of fluid bodies and the Einstein-Boltzmann equation. See Rendall (2008) for further discussion.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

(see, e.g., Ringström 2009). This apparatus implies a direct response to the hole argument. What these arguments establish is often called 'geometric uniqueness' in the mathematical physics literature, to indicate that an equivalence class of spacetimes has the 'same geometry'. But it is clear from the way that existence and uniqueness results are formulated that we cannot even pose a threat to determinism without assuming some further structure. These results establish the existence of embedding maps between manifolds (an abstract 'initial data' space and solutions to EFE), and in this context the solutions are only individuated up to isomorphism. There are no grounds to draw a further distinction between 'hole diffeomorphs'.

Applied mathematicians usually impose a third condition to have a well-posed IVF: the solutions must depend continuously, in a sense to be made precise, on the initial data.[17] Philosophers have for the most part not equated determinism with the existence of a well-posed IVF because they reject this third condition. Several alleged failures of determinism depend on exact specification of a set of initial conditions. In Xia's 'space invaders' example mentioned above, the motion of the five particles has to be choreographed with exquisite precision such that they all fly off to infinity. More precisely, the set of initial conditions that lead to the desired behavior is Lebesgue measure zero in the space of initial data (Xia 1992). Initial conditions 'arbitrarily close' to those leading to the space invaders trajectory have quite different dynamical behavior, violating the continuity requirement.[18]

This third condition raises a number of more subtle questions than the requirements of existence and uniqueness. In particular, the assessment of continuity and stability depends on the topology assigned to the space of initial data and the solution space. Yet the grounds for choosing the topology are not always clear. Fletcher (2016) considers stability questions regarding full spacetimes rather than initial data—that is, in terms of a topology on the space of Lorentzian metrics over a fixed manifold. He argues that no proposed candidate for a 'canonical' topology on this space yields intuitively correct verdicts on convergence of sequences, continuity of families of metrics, and the stability or genericity of global properties. He advocates instead a version of reflective equilibrium, according to which the choice of topology is guided by contextual features of specific applications. For the initial

17. A map $\gamma : \mathcal{X} \to \Gamma$ between initial data $\mathcal{X}$ and solutions $\Gamma$ is continuous if the inverse images under $\gamma$ of all open sets in $\Gamma$ are open sets in $\mathcal{X}$. Intuitively, points that are nearby in $\mathcal{X}$ are mapped into nearby points in $\Gamma$, such that small changes in initial data lead to small changes in the solution.

18. Xia obtains the initial conditions through a Cantor-set-like construction: for each possible interaction among the particles, eliminate the initial conditions that do not lead to the desired trajectory. Iterating this process leads to a measure zero set of initial conditions, with trajectories confined to the unstable submanifold corresponding to space invaders behavior.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

value problem, the context requires a characterization of the initial data and a choice regarding what counts as a solution to the PDE.[19] These decisions may limit the choice of topology in specific applications, perhaps even leaving us with only one viable candidate, although the physical significance of the topology and the justification for these choices deserves further scrutiny. Despite these open foundational questions, the third requirement strikes us as a plausible candidate for a principle that is required for successful applications of differential equations. We are finite epistemic agents, unable to determine initial conditions or carry out the numerical analysis of PDEs with indefinite precision. A stability requirement is thus essential to using PDEs in the study of natural phenomena. Most philosophers, by contrast, seem to allow for indefinitely precise individuation of solutions in constructing instances of indeterminism. On this view, the existence of a well-posed IVF is sufficient for determinism, but it is not necessary. In any case, we follow the philosophical tradition below in equating determinism with only the first two conditions of an IVF—the existence and uniqueness of solutions to the relevant equations—while also acknowledging the relevance of stability considerations.

**4. Global Hyperbolicity.** The seminal result due to Choquet-Bruhat and Geroch mentioned above depended on identifying a property of spacetimes necessary to complete the proof: global hyperbolicity.[20] For such spacetimes, this result establishes that EFE admits a well-posed IVF and in that sense settles the status of determinism. But not all spacetimes are globally hyperbolic. This property is in fact the top of a hierarchy of restrictions on the global causal structure of $\langle \mathcal{M}, g_{ab} \rangle$. There are solutions of EFE in which this property and properties further down in the hierarchy fail to hold. Perhaps the most famous of these is Gödel spacetime, which does not contain a global time slice $\Sigma$ and has closed time-like curves (CTCs) passing through every point. Such spacetimes are causally so pathological that the very application of the concept of determinism remains unclear. But there are cases in which we can assess determinism when global hyperbolicity does not hold. Here we will consider briefly two very different examples, which challenge whether global hyperbolicity is necessary or sufficient for determinism in GR.

19. The initial data and solutions are typically required to be elements of Sobolev spaces (spaces of functions whose partial derivatives up to some specified order are square integrable). The relationship between the function spaces for the initial data and solutions is partially dictated by the PDEs themselves and partially by the choice of desirable properties of solutions.

20. Earlier studies led to the formulation of various conditions needed to ensure the existence of an IVF for hyperbolic PDEs; global hyperbolicity is an extension of these earlier conditions to the context of GR.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

Global hyperbolicity may not be a necessary condition for determinism. The discussion above restricted attention to the type of initial value problem natural for hyperbolic equations: initial data specified on a slice $\Sigma$, with no need to specify boundary conditions on a time-like surface. But EFE allows for quite different setups, including initial-boundary value problems with input data specified on a cylindrical surface $\Sigma \cup T$ (consisting of a space-like 'end cap' $\Sigma$ and time-like sides $T$; see fig. 1). Friedrich and Nagy (1999) formulated the initial-boundary value problem for vacuum EFE, establishing that appropriate initial data on a surface $\Sigma \cup T$ determine the solution within the cylinder. This new setting leads to significant technical complications. In particular, there is not a clean separation between the dynamical evolution of the initial data and the constraints imposed on $\Sigma$, and further conditions have to be satisfied where $\Sigma$ and $T$ meet. Establishing whether two sets of initial data generate the same 'spacetime geometry' is also more challenging. Yet this setup is particularly useful for problems in numerical relativity.[21]

It is possible to formulate a well-posed initial-boundary value problem for EFE with remarkably weak assumptions about causal structure. Anti-de Sitter (AdS) spacetime is the maximally symmetric solution with $\Lambda < 0$, often studied in string theory. The light cones 'broaden out', such that the solution fails to be globally hyperbolic. Given a slice $\Sigma$ and a point $p$ in $I^+(\Sigma)$, there are past directed causal curves from $p$ that do not intersect $\Sigma$. This suggests, just as in the example of Newtonian space invaders mentioned above, that initial data on a slice $\Sigma$ do not fix future evolution because of the possibility of a disturbance coming 'from infinity'. Despite this bizarre causal structure, there is evidence supporting the conjecture that EFE admit a well-posed initial-boundary value problem in AdS for suitable boundary data on a time-like boundary.[22] The conjecture holds that AdS is asymptotically stable for optimally dissipative boundary conditions, with maximal energy flux through the time-like boundary. (By contrast, AdS appears to be unstable for the other extreme boundary condition—i.e., a reflecting boundary with vanishing flux.)

In terms of the implications for determinism, these types of solutions lead to questions similar to those raised by the space invaders example in Newtonian physics. In the Newtonian case, do we take the equations to describe a system with a fixed number of interacting particles? The affirmative answer often assumed in applying the theory implicitly imposes boundary conditions on a time-like surface. Similarly, treating gravitational degrees of freedom in AdS as part of an initial-boundary value problem leads to a well-posed problem only if we impose further conditions on the time-like boundary. In this

21. See Reula and Sarbach (2011) for an overview.

22. See Holzegel et al. (2020) for a statement of the conjecture, discussion of evidence for asymptotical stability of AdS to linearized perturbations with dissipative boundary conditions, and further references.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

case, the appropriate boundary conditions require that the system constantly "leaks" energy associated with the gravitational degrees of freedom, preventing the amplification of perturbations reflecting from the boundary. Obviously this boundary condition does not itself follow from the dynamics but has to be, in effect, imposed by hand as part of setting up the problem. The result shows that, at least in this specific case, the threat to deterministic evolution left open by the failure of global hyperbolicity can be countered by a global stipulation with a strikingly different character. We do not know whether other nonglobally hyperbolic spacetimes also admit well-posed initial-boundary value problems. In any case, although we will not pursue the point in detail here, this case raises similar questions to those we discuss below regarding global hyperbolicity.

Let us turn now to the question of whether global hyperbolicity is sufficient to guarantee a form of determinism. To state the problem clearly we need to first introduce the idea of an extension: a spacetime $\langle \mathcal{M}, g_{ab} \rangle$ can be *extended* if it is isometric to a proper subset of another spacetime.[23] An extension of a spacetime is *maximal* if it is *inextendible*; that is, it is not isometric to a proper subset of another spacetime. The result cited above establishes that appropriate initial data fix a solution throughout the domain of dependence (up to isomorphism). But are they sufficient to establish global uniqueness? Suppose that the domain of dependence has a future boundary, called the *Cauchy horizon*, $H^+(\Sigma)$. What determines evolution in the regions beyond this boundary?

Taub-NUT spacetime, a solution of the vacuum EFE, illustrates a failure of global uniqueness. One region of this spacetime has global slices $\Sigma$ with a nontrivial domain of dependence (the *Taub region* of the spacetime), which is thus globally hyperbolic; however, $D(\Sigma)$ is not maximal and it admits of multiple extensions (the *NUT regions*; see fig. 2). Topologically the Taub region is $S^3 \times (t_-, t_+)$, bounded by Cauchy horizons at $t_-$ and $t_+$ ($\in \mathbb{R}$) generated by closed null geodesics. For any point in the Taub region, there are complete null geodesics that cross the horizon, but there are also incomplete null geodesics that spiral toward the Cauchy horizon. Despite this singular behavior, the Taub region admits two extensions in both directions—that is, through $H^+(\Sigma)$ and $H^-(\Sigma)$. The sense in which the existence of these extensions leads to a failure of uniqueness is surprisingly subtle, however (Chruściel and Isenberg 1993): the two extensions through either horizon are locally isometric, but combinations of the different extensions through both horizons can generate nonisometric spacetimes. To illustrate this, let

23. That is, it can be extended if there exists a spacetime $\langle \mathcal{M}', g'_{ab} \rangle$, $\mathcal{M} \subsetneq \mathcal{M}'$, and an isometric embedding $\phi : \mathcal{M} \to \mathcal{M}'$ such that $\forall\, p \in \mathcal{M}'$, $\phi^*(g_{ab}(\phi^{-1}(p))) = g'_{ab}(p)$. We can further refine this definition by specifying the differentiability class of the isometries.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

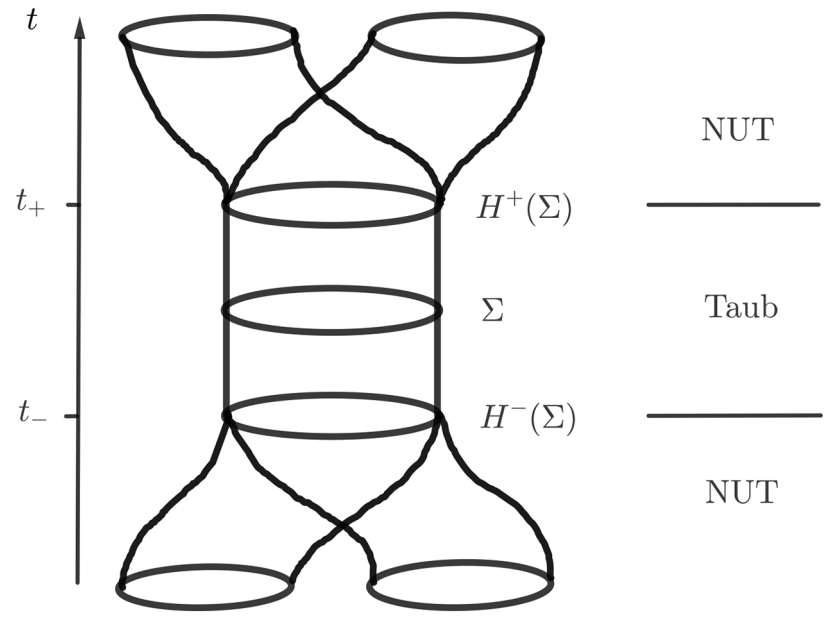



Figure 2. In Taub-NUT spacetime, the Taub region has two extensions (NUT regions) each to the future and the past. Although the figure shows all these extensions at once, we only mean to highlight the distinct possibilities and not that the spacetime is itself in some sense 'branching'.

us call the bottom two extensions in figure 2 before $t_-$ 'legs' and the top ones after $t_+$ 'arms'. The two legs are locally isometric, and so are the two arms. However, the spacetime resulting from adding, say, the left arm and the left leg is not isometric to the one resulting from adding the left arm and the right leg.

Chruściel and Isenberg (1993) generalize this result using the study of two other families of exact solutions with Cauchy horizons: there appear to be multiple nonisometric extensions in cases in which, as in Taub-NUT, the Cauchy horizon consists of disconnected components. In these cases the failure of uniqueness has a striking global character, since it only arises in considering extensions through disconnected components of the Cauchy horizon. There is further work to be done to establish that uniqueness also fails because of local extensions through a Cauchy horizon.

If such cases are deemed to be physically possible, then even though EFE locally admits a well-posed IVF this cannot be extended to prove global existence and uniqueness. We explore below various different stances on whether this simple example can be ruled out in some way. The extensions may necessarily exploit symmetries of these exact solutions, running afoul

https://doi.org/10.1086/713904 Published online by Cambridge University Press

of the requirement of stability, or perhaps the unusual causal structure of these models is objectionable. But we should emphasize here that these different responses share one feature, namely, that they introduce a definition of physical possibility that goes beyond imposing the local dynamical laws. In fact, this is necessarily so, as the Taub-NUT spacetime and similar examples are solutions of EFE.

We now move to the discussion of the three approaches of saving determinism in GR, focusing on global hyperbolicity as an at least necessary condition for determinism. These three approaches conceive of the status of global hyperbolicity as a condition (i) to be stipulated a priori, (ii) to be proved from weaker assumptions within GR, and (iii) to be confirmed—or infirmed—from physics beyond GR.

**5. Global Hyperbolicity by Fiat.** Global hyperbolicity is the primary condition necessary to ensure determinism in GR (modulo the concerns in the previous section). The first approach to explicate the status of the global condition of global hyperbolicity essentially stipulates its validity. It is based on a transcendental argument in that its imposition by fiat acts as a condition for obtaining a sensible physical theory. This kind of argument is not without precedent: we routinely ask our theories to satisfy certain conditions dictated by the mathematics. For instance, in classical mechanics, the Lipschitz condition is imposed in order to satisfy an antecedent condition of important existence and uniqueness theorems for the PDEs capturing the standard IVF. Thus, its violation leads to forms of indeterminism as exemplified in Norton's dome.

In spacetime physics, one routinely requires that the manifold be a *Hausdorff space*, that is, a topological space such that any two distinct points have disjoint neighborhoods.[24] Typically, demanding that a spacetime manifold be a Hausdorff space is not considered a stipulation with all the advantages of theft over honest toil but instead a refinement of what is needed mathematically to formulate a sensible theory. Penrose (1979, 595) affirms this when he writes that "I must . . . return firmly to sanity by repeating to myself three times: 'spacetime is a *Hausdorff* differentiable manifold; spacetime is a *Hausdorff*. . . .'" The need to refine our definition of a manifold only became apparent with the recognition of bizarre non-Hausdorff cases, for example, the cases of branching spacetimes considered, but rejected, by Penrose. Although explicitly defended, for example, in Belnap (1992) and McCall (1994), Earman (2008) argues that actually branching individual

24. There are other routine requirements with a similar character, e.g., that spacetime must be "as large as it can be" or free of holes. We do not have space to consider hole-freeness conditions in detail here, but see Manchak (2016) and Doboszewski (2020) for excellent recent discussions.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

spacetimes are not necessary to model indeterminism. We share Earman's viewpoint: our characterization of (in)determinism does not require actual branching of spacetimes. Furthermore, and more to the point, Earman (2008, sec. 3) shows that important theorems assume the Hausdorff property. The theorems listed by Earman exemplify the sense in which the Hausdorff property is considered part of the mathematical machinery necessary for a sensible formulation of GR. Although advocates of branching spacetimes have responded by articulating how appropriate generalizations of the notion of manifolds circumvent such problems to some degree (Placek and Belnap 2012; Müller 2013; Luc and Placek 2020), it ought to be noted that the non-Hausdorff manifolds are deployed by these authors to fulfill a rather different representational task: the non-Hausdorff structure represents a space of 'possible spacetimes' all crowded into a single manifold.

With the standard representational duties in GR in mind, the Hausdorff property is thus routinely stipulated or implicitly assumed in order to guarantee the smooth workings of the mathematical apparatus. The idea behind taking the status of global hyperbolicity to be that of an a priori demand on candidate spacetime models in GR is rather similar: all globally hyperbolic spacetimes are straightforwardly amenable to a well-posed IVF.

The problem with this approach is that global hyperbolicity is not obviously analogous to the Hausdorff condition, as it turns out that one can formulate a sensible theory in nonglobally hyperbolic spacetimes (see Friedman 2004). While the latter is widely acknowledged to be a mathematical condition that imposes no physically relevant constraint, global hyperbolicity is thus a substantive assumption. Even if one thinks that determinism is a central physical principle without which no physical theory can operate, enjoining global hyperbolicity thus requires a similarly substantive argument why we ought to accept its a priori status. An immediate objection to any resistance to demanding global hyperbolicity would be to question the relevance of nonglobally hyperbolic spacetimes, for example, spacetimes with CTCs in their maximal extension. Given that we take our universe to have the kinds of features it has, we would not expect that it is best described by a spacetime with CTCs.[25] The problem with this objection is that if successful, it would only imply that the actual world is deterministic. It would not, however, have any implications as to whether the theory—GR—is deterministic. To the extent to which we are interested in a question about GR, we have to consider the full space of solutions, rather than just some particular solutions that happen to be useful in some particular situations. But this means that we

25. A possible exception to this are Kerr-Newman spacetimes, which contain CTCs in their maximal analytic extension; they describe rotating, charged black holes, which are believed to be present in many regions of spacetime.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

should not simply accept global hyperbolicity as an additional condition that is tacked onto GR in the same way that we ask our manifolds to be Hausdorff spaces.

Now it seems odd to ask whether we could establish, by observation, that spacetime is Hausdorff. However, one may reasonably hope to justify global hyperbolicity, just because it is a physically substantive assumption, not by fiat but through respectable a posteriori observation. The well-known problem is that observations from the confines of a spacetime point, or even a compact region, underdetermine the cosmological model that best describes our actual universe (see Glymour 1977; Malament 1977). Recently, Manchak (2009) has proved that, barring one type of pathology, every spacetime is observationally indistinguishable from some other spacetime with different global structure, even fixing all the local properties of the spacetime such as the validity of EFE. The exception includes spacetimes in which there is a point $p \in \mathcal{M}$ such that $I^-(p) = \mathcal{M}$. Not only can all of the spacetimes be surveyed from one point in this case, but it can be shown that these spacetimes all contain CTCs. Since spacetimes of this class are thus not globally hyperbolic, this exception is not going to do any work for the defender of a posteriori global hyperbolicity: either the spacetimes are already non-globally hyperbolic or else their global structure is underdetermined.

This result in itself does not completely spoil this line of reasoning. For even though the global structure of spacetime is generically underdetermined given $I^-(p)$ for any $p$, it could still be the case that partial determinations of global structure can be made. In particular, it could be hoped that the global hyperbolicity can be observationally established, at least in principle, while the complete global structure cannot. Alas, Manchak (2011) proves that we cannot even make, locally, the determination of whether spacetime is globally hyperbolic.[26] In his interpretation, this means that although our universe may in fact be globally hyperbolic, we cannot know that it is.[27] It should not strike us as particularly surprising that it cannot be determined, from causal information available at a single point, whether spacetime is globally hyperbolic, that is, effectively whether the causal signals emanating from everywhere in spacetime must all intersect a global time slice.[28] Global hyperbolicity seems to be a property of spacetimes squarely at odds with the project of determining what constrains the global spacetime structure, given our epistemic limitations.

26. This theorem also excludes the same pathological cases.

27. If our universe fails to be globally hyperbolic, however, then we may know that (Malament 1977).

28. Assuming that causal signals propagate along causal curves and that these curves exemplify the necessary conditions of inextendibility.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

Thus, global hyperbolicity cannot legitimately be established by fiat and not possibly by observation. Consequently, it assumes the status of a substantive open issue that must be attacked with all the resources the theory can offer. In fact, the *(strong) cosmic censorship hypothesis*, roughly the hypothesis that all physically reasonable spacetimes are globally hyperbolic, is widely held to be the most important open issue in the foundations of GR. Attempts to prove censorship theorems from the resources of GR, and thus to show that global hyperbolicity can be inferred from weaker assumptions, are the topic of the next section.

**6. Censorship Theorems.** Hawking motivates censorship theorems as follows (Hawking and Penrose 1996, 10): "Global hyperbolicity may be a necessity. But my viewpoint is that one shouldn't assume it because that may be ruling out something that gravity is trying to tell us. Rather, one should deduce that certain regions of spacetime are globally hyperbolic from other physically reasonable assumptions." The input assumptions include constraints on the matter sources plugged into EFE and a requirement that the initial data are appropriately 'generic'; a proof would show that such initial data yield globally hyperbolic maximal developments. The name derives from the idea that gravity would act as a 'censor', effectively preventing failures of causality from occurring or (in a weaker version of the hypothesis) keeping failures of causality hidden behind the event horizons of black holes. Given a successful proof, gravity would be telling us that the local existence and uniqueness theorems for GR can be extended globally, ruling out multiple extensions such as the example from section 2. In more physical terms, these results would vindicate Penrose's seminal idea that physical effects near a Cauchy horizon generate curvature singularities, preventing extensions.

Although it is difficult to formulate the input assumptions precisely, they are arguably well motivated by the desire to isolate dynamical effects of GR. There are two distinct issues. First, one limits consideration to 'suitable' matter fields—those that do not themselves develop singularities in Minkowski spacetime. The idea is to separate singularities due to the matter fields themselves from those due to gravity; it rules out, for example, certain types of fluids because they develop singular shock waves. The second qualification allows that there may be cases based on highly 'special' initial data sets (e.g., with a high degree of symmetry) that lead to a failure of uniqueness.

This leads to the following formulation of strong cosmic censorship (SCC; Chruściel 1991). Given (i) a space of initial data $\mathcal{X}(\Sigma, h_{ab}, k_{ab})$ satisfying the Gauss-Codacci constraints, supplemented with appropriate initial data for matter fields, and (ii) a collection of four-dimensional Lorentzian manifolds $\mathcal{S}(M, g_{ab})$, where the metric $g_{ab}$ has some specified differentiability class, SCC then states that there is a set $\mathcal{G}(\Sigma, h_{ab}, k_{ab})$ in $\mathcal{X}(\Sigma, h_{ab}, k_{ab})$ such

https://doi.org/10.1086/713904 Published online by Cambridge University Press

that the MGHD for every data set in $\mathcal{G}$ is inextendible in $\mathcal{S}(M, g_{ab})$. Set $\mathcal{G}$ is a proper subset of $\mathcal{X}$, as one needs to impose some further conditions in addition to the constraints to ensure that the initial data are themselves 'maximal' in an appropriate sense.[29] We further require that $\mathcal{G}$ must be an open, dense set in an appropriate topology. This reflects the stability physicists usually require for a well-posed IVF, mentioned at the end of section 3, but faces the same challenges. Specifying what qualifies as 'special' or 'generic' initial conditions requires introducing a measure or at least a topology on the space of initial data, but as Wald (1998) emphasizes, there is not currently a standard measure or topology on the space of initial data, or solutions, that we can appeal to.[30]

There are several purported counterexamples to SCC in the literature, in the form of initial data with MGHD that admit nonisometric extensions given a choice of $\mathcal{S}(M, g_{ab})$. Yet there are active debates regarding the implications and the appropriate response to these results.

Regarding our example from section 2, we need not worry about singularities arising from matter fields because Taub-NUT is a vacuum solution. However, stability considerations would presumably rule it out as a counterexample due to the symmetry of the spacetime. There are some results indicating that Cauchy horizons are 'special' in the sense of being measure zero in the space of solutions of EFE. For example, Moncrief and Isenberg (1983) prove that in a particular case (i.e., granted that the Cauchy horizon is compact, analytic, and ruled by closed null geodesics) there are symmetries in the neighborhood of the Cauchy horizon. As a result, this family of solutions would not be expected to fill an open set of initial data in any reasonable topology. Moncrief and Isenberg (2020) have extended their results regarding the existence of symmetries to cover a broader class of spacetimes with Cauchy horizons. General results along these lines support SCC by showing that the existence of Cauchy horizons does not hold generically, for open sets of initial data, rather than just for specific cases.

Alongside work showing that purported counterexamples fail the stability requirement, there are a number of theorems establishing global uniqueness for specific choices of $\mathcal{G}$. In the best case, a proof of SCC would follow for all initial data satisfying modest requirements, such as completeness and falloff conditions. Current results fall short of that aim but show that the SCC holds for several quite different families of models. For example, Christodoulou

29. These must at least include a requirement that the initial data are complete or satisfy 'falloff' conditions (such as asymptotic flatness); otherwise, initial data specified on a compact region on a space-like hypersurface in Minkowski spacetime provide a counterexample. See, e.g., Ringström (2009) for further discussion.

30. The challenges are similar to those Fletcher (2016) identifies with finding a topology on the space of spacetimes.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

and Klainerman (1993) established the stability of flat spacetime to small perturbations; roughly, for initial data 'close to' Minkowski spacetime, with appropriate decay at spatial infinity, $D(\Sigma)$ is inextendible. Studies of polarized Gowdy spacetimes (a family of symmetric spacetimes with two independent space-like Killing fields) have shown that, generically, curvature invariants blow up along curves approaching the Cauchy horizon. These results establish the existence of a singularity that prevents extensions through the horizon.[31] Ringström (2009, pt. 4) considers the fate of the SCC within the Bianchi models and shows that they fall into two classes: the first exhibits curvature singularities, so that the MGHD cannot be extended, whereas the second class does admit nonisometric extensions of the MGHD. Yet Ringström further shows that all of the models in this second case have a further symmetry, suggesting that they would fail to satisfy the stability requirement. In sum, present results constitute a patchwork, showing that the SCC holds for a variety of specific choices of initial data but falling short of a fully general result.

A different set of issues arises with regard to the choice of $\mathcal{S}(M, g_{ab})$: in particular, should this set include Lorentzian manifolds where $g_{ab}$ is continuous ($C^0$) but lacks well-defined derivatives? Several purported counterexamples to SCC assume a positive answer. Dafermos and Luk (2017), for example, construct a continuous extension through the Cauchy horizon in Kerr solutions. Yet the physical significance of merely $C^0$ metrics is not clear. Second derivatives of the metric appear in the Einstein tensor in the EFE, suggesting that we should impose stronger requirements—most straightforwardly, that $g_{ab}$ is at least $C^2$—but slightly weaker constraints may allow for EFE to be well defined in a distributional sense at the Cauchy horizon (Earman 1995, 46–50).

Furthermore, it should be noted that the choice of $\mathcal{S}(M, g_{ab})$ has implications for the very definition of what it means for a MGHD to be 'extendible'.[32] The property of 'inextendibility' depends on what one takes the set of relevant models to be: one and the same Lorentzian manifold may be inextendible with respect to one background collection of models but extendible with respect to another. Restrictions to 'physically reasonable' models, $\mathcal{P} \subset \mathcal{S}$, affect assessments of modality. Manchak (2018) argues that, for every nontrivial choice of $\mathcal{P}$, it is not the case that every $\mathcal{P}$-inextendible $\mathcal{P}$-model is inextendible in $\mathcal{S}$. This obviously has a direct bearing on debates regarding the SCC, given that alleged counterexamples may fall outside some chosen collections $\mathcal{P}$.

Decades of effort have led to much sharper formulations of the SCC, yet the proof of a general version of the SCC has remained elusive. But perhaps

31. See Isenberg (2015) for further discussion of Gowdy spacetimes and Andersson (2004) for a more general review.

32. We thank a referee for pointing this out to us.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

the results that have already been achieved are sufficient; in what sense would a patchwork quilt of stability results answer our original question regarding whether GR is deterministic? In the best case scenario, the sets $\mathcal{G}$ for which SCC holds would cover all the solutions used in black hole physics or cosmological modeling. This would involve a much more parochial sense of physical possibility, as it would apply directly to only a subset of the full solution space of GR. Although the question of whether classical GR is deterministic, full stop, would remain unresolved, such results would provide assurance that no failure of determinism would crop up in the applications of GR to physical systems covered by results making up the quilt.

**7. Incompleteness of GR.** Nobody expects GR to be the last word on gravitation. Accordingly, an advocate of global hyperbolicity may draw hope from the fact that nonglobally hyperbolic spacetimes may be incompatible with the quantum. As it turns out, however, the direction the physics literature indicates in this respect is far from unambiguous. For instance, considerations in semiclassical quantum gravity (QG) pull both ways. On the one hand, the usual point-wise energy conditions all fail for quantum fields. Thus, quantum effects may undermine efforts to establish cosmic censorship from assumptions concerning what energy conditions the energy-matter content of the universe may reasonably be expected to satisfy. On the other hand, quantum effects may effectively act as barriers at Cauchy horizons, cutting off spacetime extensions beyond the globally hyperbolic domain.[33]

While results in semiclassical QG pull both ways, an attempt to formulate a QFT on curved spacetimes may provide respectable grounds to stipulate global hyperbolicity from the get-go. A standard recipe for obtaining a natural algebra $\mathfrak{A}(\mathcal{M}, g_{ab})$ of observables for a quantum field defined over a spacetime $\langle \mathcal{M}, g_{ab} \rangle$ only succeeds if the spacetime is globally hyperbolic. But the defender of global hyperbolicity will not find much solace in this, for two reasons. First, since QFT on curved background spacetimes does not consider the back reaction of the quantum fields on the spacetime it has no prayer of being a fundamental theory; at best, it is a 0th approximation to a full quantum theory of gravity that can be useful to model the weak-field regime. Second, it seems possible that the conditions necessary for the construction of the algebra of observables may be weakened in such ways as to enable the procedure to apply to nonglobally hyperbolic spacetimes. To be sure, some of these relaxations are problematic, but it is far from clear how sacrosanct the condition of global hyperbolicity really is.[34]

33. For a survey of related results, see Earman et al. (2009, sec. 5) and Smeenk and Wüthrich (2011, sec. 7).

34. For further details, see Earman et al. (2009, sec. 6).

https://doi.org/10.1086/713904 Published online by Cambridge University Press

Of course, only a full theory of QG will be able to give more authoritative answers. But since there exist such a plethora of competing approaches, it is currently impossible to draw many general conclusions from research in QG. One possibility that must be taken into serious consideration, however, is the suggestion offered, in one form or another, by many approaches to QG that spacetime fails to be fundamental but instead emerges from an underlying, nonspatiotemporal structure (Huggett and Wüthrich 2013). In this case, it may become impossible to straightforwardly connect global hyperbolicity, in the sense defined above, to this fundamental structure. Assuming, however, that this can be done—as it has to in order for the theory to have a meaningful classical limit—we see in principle three different manners in which the issue of global hyperbolicity may play out in QG. First, it may be postulated by fiat (as discussed in sec. 5) as a transcendental condition for the approach. This is unproblematic as long as it is understood that the assumption ultimately requires either independent support or else justification ex post facto from the success of the theory based on it. Second, global hyperbolicity, or what amounts to it in the appropriate classical limit, may be derived from the resources of the approach itself or from conditions solidly grounded in considerations concerning what is physically reasonable drawing, perhaps, on the resources of other theories. As a third possibility, it may turn out that global hyperbolicity or its quantum analogue is not vindicated by the fundamental theory. In this latter case, it is hard to imagine a sound reason for insisting on global hyperbolicity at the classical level.

The third possibility could be borne out in string theory. There are strong indications that at least in five-dimensional supersymmetric gravity, a close relative to string theory, analogues of spacetimes with CTCs exist, and thus violations of global hyperbolicity may occur (Gauntlett et al. 2003). Since solutions of the five-dimensional theory can easily be worked up to solutions for ten- and eleven-dimensional supergravity, the situation in string theory is arguably not much different from the one in GR. The second option above would amount to an attempt to prove an analogue of cosmic censorship in full QG. Finally, several approaches, among them causal sets and loop quantum gravity (LQG), go for the first alternative; that is, they essentially impose global hyperbolicity or a close cousin ab initio.[35] The canonical quantization procedure applied in LQG simply requires global hyperbolicity in order to get traction. However, an alternative viewpoint conceives of LQG not as primarily a theory concerned with the global structure of spacetime but rather as offering a quantum theory of small patches of spacetime as may be measured in a laboratory. This alternative would view LQG not as

35. Although it is not obvious what the discrete analogue of global hyperbolicity would be, it seems natural to think that an adequately transposed condition holds in the causal sets approach. Its basic assumption of acyclicity certainly rules out CTCs.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

cosmological but instead as concerned with local physics (Wüthrich, forthcoming). We believe that it is unfortunate that this viewpoint is rarely taken seriously, but we submit that, according to it, LQG could not be a fundamental quantum theory of spacetime capable of accounting for the global structure of spacetime.

As is evident, once again, the field is simply too rich to draw robust general conclusions concerning the fate of determinism. But even though this may come as a disappointment to some, we believe that the many foundational lessons that can be gleaned from the field make the endeavor more than worthwhile.

**8. Conclusions.** In our investigation of the fate of determinism in GR and beyond, we have explored three different approaches to justifying the imposition of global hyperbolicity. The first approach aimed to justify this condition directly. Global hyperbolicity is a global property of spacetime, and it generally cannot be established via observations of a local region. So it appears that the condition has to be treated as an a priori constraint, and we considered, but rejected, the possibility of treating it as a mathematical constraint transcendentally necessary for the formulation of a sensible theory. The second approach seeks to establish 'censorship theorems', showing that global hyperbolicity holds for some subset of spacetimes that satisfy weaker, physically motivated requirements. While there are some promising results, this approach is a work in progress, and it remains genuinely open whether it will succeed in establishing global hyperbolicity, at least for some relevant classes of models. Finally, a third approach insists on the need to consider the limits of applicability of classical GR and its interconnections with other theories, such as QFT on curved spacetimes and semiclassical and full QG. Although more speculative, this approach bears the promise to elucidate the relationship between GR and more fundamental theories toward QG and thus may have significant heuristic value for articulating a theory of QG.

Our discussion of the fate of determinism in GR required a decision regarding the concept of determinism itself. The standard Laplacian notion according to which the state of the world at a time together with the laws of nature jointly determine the state of the world at all times naturally translates into the requirement that the (geometric and matter) data on a global time slice $\Sigma$ uniquely extend to all of spacetime. The existence and uniqueness of such an extension thus required for the satisfaction of determinism as usually envisioned by philosophers does not, however, suffice for a well-posed IVF as sought after by physicists. In addition, solutions must continuously depend on the initial data. Although technically involved, this continuity condition, if added, rules out 'unphysical' failures of determinism due to exquisitely fine-tuned initial data. Thus, the question of determinism in GR not only raises deep and fruitful philosophical and foundational issues, but it also

https://doi.org/10.1086/713904 Published online by Cambridge University Press

serves to bring out a contrast between philosophical interests and the demands of applied mathematics.

We hope to have highlighted, once again, the value of assessing the fate of determinism in modern physics. There is simply no quick answer to be had, as it depends on a plethora of subtle conceptual and technical issues in the foundations of the theories considered. But this is what makes determinism such a central topic in the exploration of physical theories in general and of GR in particular.

## REFERENCES


Andersson, Lars. 2004. "The Global Existence Problem in General Relativity." In *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, ed. Piotr Chruściel and Helmut Friedrich, 71–120. Boston: Birkhäuser.

Belnap, Nuel. 1992. "Branching SpaceTime." *Synthese* 92:385–434.

Budic, Robert, and Rainer K. Sachs. 1976. "Deterministic SpaceTimes." *General Relativity and Gravitation* 7:21–29.

Butterfield, Jeremy. 1989. "The Hole Truth." *British Journal for the Philosophy of Science* 40:169–88.

Choquet-Bruhat, Yvonne, and Robert Geroch. 1969. "Global Aspects of the Cauchy Problem in General Relativity." *Communications in Mathematical Physics* 14:329–35.

Christodoulou, Demetrios. 2007. *The Formation of Shocks in 3-Dimensional Fluids*, vol. 2. Zurich: European Mathematical Society.

Christodoulou, Demetrios, and Sergiu Klainerman. 1993. *The Global Nonlinear Stability of the Minkowski Space*. Princeton, NJ: Princeton University Press.

Chruściel, Piotr T. 1991. *On Uniqueness in the Large of Solutions of Einstein's Equations ("Strong Cosmic Censorship")*. Proceedings of the Centre for Mathematics and Its Applications, Australian National University 27. Canberra: Australian National University.

Chruściel, Piotr T., and James Isenberg. 1993. "Nonisometric Vacuum Extensions of Vacuum Maximal Globally Hyperbolic Spacetimes." *Physical Review* D 48 (4): 1616–28.

Dafermos, Mihalis, and Jonathan Luk. 2017. "The Interior of Dynamical Vacuum Black Holes I: The $C^0$-Stability of the Kerr Cauchy Horizon." arXiv, Cornell University. https://arxiv.org/abs/1710.01722.

Doboszewski, Juliusz. 2019. "Relativistic Spacetimes and Definitions of Determinism." *European Journal for Philosophy of Science* 9:24.

———. 2020. "Epistemic Holes and Determinism in Classical General Relativity." *British Journal for the Philosophy of Science* 71 (3): 1093–111.

Earman, John. 1986.*A Primer on Determinism*. Dordrecht: Kluwer.

———. 1995. *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. New York: Oxford University Press.

———. 2004. "Determinism: What We Have Learned and What We Still Don't Know." In *Determinism, Freedom, and Agency*, ed. Joseph K. Campbell, Michael O'Rourke, and David Shier. Cambridge, MA: MIT Press.

———. 2008. "Pruning Some Branches from 'Branching Spacetimes.'" In *The Ontology of Spacetime II*, ed. Dennis Dieks, 187–206. Amsterdam: Elsevier.

Earman, John, and John Norton. 1987. "What Price Spacetime Substantivalism? The Hole Story." *British Journal for the Philosophy of Science* 38:515–25.

Earman, John, Chris Smeenk, and Christian Wüthrich. 2009. "Do the Laws of Physics Forbid the Operation of Time Machines?" *Synthese* 169:91–124.

Fletcher, Samuel C. 2012. "What Counts as a Newtonian System? The View from Norton's Dome." *European Journal for Philosophy of Science* 2:275–97.

———. 2016. "Similarity, Topology, and Physical Significance in Relativity Theory." *British Journal for the Philosophy of Science* 67:365–89.


https://doi.org/10.1086/713904 Published online by Cambridge University Press

Friedman, John L. 2004. "The Cauchy Problem on Spacetimes That Are Not Globally Hyperbolic." In *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, ed. Piotr Chruściel and Helmut Friedrich, 331–46. Boston: Birkhäuser.
Friedrich, Helmut, and Gabriel Nagy. 1999. "The Initial Boundary Value Problem for Einstein's Vacuum Field Equation." *Communications in Mathematical Physics* 201:619–55.
Friedrich, Helmut, and Alan Rendall. 1999. "The Cauchy Problem for the Einstein Equations." In *Einstein's Field Equations and Their Physical Implications*, ed. Bernd G. Schmidt, 127–223. Lecture Notes in Physics. Dordrecht: Springer.
Gauntlett, Jerome P., Jan B. Gutowski, Christopher M. Hull, Stathis Pakis, and Harvey S. Reall. 2003. "All Supersymmetric Solutions of Minimal Supergravity in Five Dimensions." *Classical and Quantum Gravity* 20:4587–634.
Geroch, Robert. 1977. "Prediction in General Relativity." In *Foundations of SpaceTime Theories*, ed. John Earman, Clark Glymour, and John Stachel, 81–93. Minneapolis: University of Minnesota Press.
———. 1996. "Partial Differential Equations of Physics." In *General Relativity: Proceedings of the Forty-Sixth Scottish Universities Summer School in Physics, Aberdeen*, ed. Graham S. Hall and John R. Pulham, 19–60. Edinburgh: SUSSP.
Glymour, Clark. 1977. "Indistinguishable SpaceTimes and the Fundamental Group." In *Foundations of SpaceTime Theories*, ed. John Earman, Clark Glymour, and John Stachel, 50–59. Minneapolis: University of Minnesota Press.
Hawking, Stephen W., and George F. R. Ellis. 1973. *The Large Scale Structure of SpaceTime*. Cambridge: Cambridge University Press.
Hawking, Stephen W., and Roger Penrose. 1996. *The Nature of Space and Time*. Princeton, NJ: Princeton University Press.
Holzegel, Gustav, Jonathan Luk, Jacques Smulevici, and Claude Warnick. 2020. "Asymptotic Properties of Linear Field Equations in Anto-De Sitter Space." *Communications in Mathematical Physics* 374:1125–78.
Huggett, Nick, and Christian Wüthrich. 2013. "Emergent Spacetime and Empirical (In)coherence." *Studies in the History and Philosophy of Modern Physics* 44:276–85.
Isenberg, James. 2015. "On Strong Cosmic Censorship." *Surveys in Differential Geometry* 20 (1): 17–36.
Luc, Joanna, and Tomasz Placek. 2020. "Interpreting Non-Hausdorff (Generalized) Manifolds in General Relativity." *Philosophy of Science* 87 (1): 21–42.
Malament, David B. 1977. "Obervationally Indistinguishable SpaceTimes." In *Foundations of SpaceTime Theories*, ed. John Earman, Clark Glymour, and John Stachel, 61–80. Minneapolis: University of Minnesota Press.
———. 2008. "Norton's Slippery Slope." *Philosophy of Science* 75:799–816.
———. 2012. *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*. Chicago: University of Chicago Press.
Manchak, John B. 2009. "Can We Know the Global Structure of Spacetime?" *Studies in History and Philosophy of Modern Physics* 40:53–56.
———. 2011. "What Is a Physically Reasonable SpaceTime?" *Philosophy of Science* 42:74–76.
———. 2016. "Epistemic 'Holes' in SpaceTime." *Philosophy of Science* 83 (2): 265–76.
———. 2018. "Some 'No-Hole' Spacetime Properties Are Unstable." *Foundations of Physics* 48:1539–45.
McCall, Storrs. 1994. *A Model of the Universe*. Oxford: Clarendon.
Moncrief, Vincent, and James Isenberg. 1983. "Symmetries of Cosmological Cauchy Horizons." *Communications in Mathematical Physics* 89:387–413.
———. 2020. "Symmetries of Cosmological Cauchy Horizons with Non-closed Orbits." *Communications in Mathematical Physics* 374 (1): 145–86.
Müller, Thomas. 2013. "A Generalized Manifold Topology for Branching SpaceTimes." *Philosophy of Science* 80:1089–100.
Norton, John D. 2003. "Causation as Folk Science." *Philosophers' Imprint* 3:1–22.
Penrose, Roger. 1979. "Singularities and Time-Asymmetry." In *General Relativity: An Einstein Centenary Survey*, ed. Stephen W. Hawking and Werner Israel, 581–638. Cambridge: Cambridge University Press.

https://doi.org/10.1086/713904 Published online by Cambridge University Press

Placek, Tomasz, and Nuel Belnap. 2012. "Indeterminism Is a Modal Notion: Branching Spacetimes and Earman's Pruning." *Synthese* 187:441–69.
Rendall, Alan D. 2008. *Partial Differential Equations in General Relativity*, vol. 16 of *Oxford Graduate Texts in Mathematics*. Oxford: Oxford University Press.
Reula, Oscar, and Olivier Sarbach. 2011. "The Initial-Boundary Value Problem in General Relativity." *International Journal of Modern Physics* D 20:767–83.
Ringström, Hans. 2009. *The Cauchy Problem in General Relativity*. Zurich: European Mathematical Society.
Sarbach, Olivier, and Manuel Tiglio. 2012. "Continuum and Discrete Initial-Boundary Value Problems and Einstein's Field Equations." *Living Reviews in Relativity* 15 (1): 9.
Smeenk, Chris, and Christian Wüthrich. 2011. "Time Travel and Time Machines." In *The Oxford Handbook of Time*, ed. Craig Callender, 577–630. Oxford: Oxford University Press.
Stachel, John. 2014. "The Hole Argument and Some Physical and Philosophical Implications." *Living Reviews in Relativity* 17 (1).
Wald, Robert M. 1998. "Gravitational Collapse and Cosmic Censorship." In *Black Holes, Gravitational Radiation and the Universe: Essays in Honor of C. V. Vishveshwara*, ed. Bala R. Iyer and Biplab Bhawal, 69–85. Dordrecht: Kluwer.
Weatherall, James O. 2018. "Regarding the 'Hole Argument.'" *British Journal for the Philosophy of Science* 69:329–50.
Wüthrich, Christian. 2011. "Can the World Be Shown to Be Indeterministic after All?" In *Probabilities in Physics*, ed. Claus Beisbart and Stephan Hartmann, 365–89. Oxford: Oxford University Press.
———. Forthcoming. "Time Travelling in Emergent Spacetime." In *Hajnal Andréka and István Németi on the Unity of Science: From Computing to Relativity Theory through Algebraic Logic*, ed. Judit Madarász and Gergely Székely. Dordrecht: Springer.
Xia, Zhihong. 1992. "The Existence of Noncollision Singularities in Newtonian Systems." *Annals of Mathematics* 135:411–68.

https://doi.org/10.1086/713904 Published online by Cambridge University Press