# Closing the Hole Argument

Hans Halvorson and JB Manchak

November 3, 2021

**Abstract**

The hole argument purportedly shows that spacetime substantivalism implies a pernicious form of indeterminism. Here we attempt to answer the question: what is the mathematical fact that is supposed to underwrite the hole argument? We identify two relevant mathematical claims. The first claim is trivially true, and Weatherall (2018) has convincingly argued that it cannot support the hole argument. The second claim would support the hole argument, but we prove that it is false. Therefore, there is no basis for the hole argument.

## 1 Introduction

In terms of generating discussion, few articles in the philosophy of physics can parallel John Earman and John Norton's (1987) article on the "hole argument" for local spacetime theories. Recall that the early twentieth century saw a revival of relationalist views of space and time, where the latter are conceived of as non-fundamental entities whose existence depends on that of their material contents. It is not surprising, then, that when metaphysical and scientific realism came back into vogue in the late twentieth century, then so did spacetime substantivalism. One might fairly say that by the 1980s, the dominant view was that four-dimensional spacetime is an independently existing substance.

Then, in a judo like maneuver, Earman and Norton (1987) put forward the hole argument, according to which a substantivalist must be committed to a pernicious form of indeterminism.[1] Earman and Norton's argument appears to be based on a subtle mathematical fact about gauge and general covariance — a fact that even Einstein had been confused about. In one fell swoop, Earman and Norton tilted the balance of power back towards relationalism, and they inaugurated a new golden-age in the philosophy of spacetime physics.

In this paper we argue that the hole argument seems plausible only because it confuses two distinct mathematical claims. The first of these claims — viz. that there are distinct but isomorphic models — is trivially true, and Weatherall (2018) has convincingly argued that it does not support the hole argument. However, some philosophers may have confused this mathematical triviality with another mathematical claim, which, if true, would imply that General Relativity is indeterministic. We prove that this second claim is false, and we conclude that there is no mathematical fact that could underwrite the hole argument.

## 2 Preliminaries

In this section we accomplish two things. First, we review the general structure of the hole argument. Second, we clarify some basic mathematical facts that are relevant for an evaluation of the hole argument. Most importantly, we clarify the distinction between diffeomorphism and isometry, and what is invariant under each of these kinds of mappings — since a failure to understand this distinction has led to many confusions regarding the upshot of the hole argument.

The hole argument has the following general structure:

(A) Substantivalism: spacetime exists, independently of the contents inside it.

(B) Some mathematical fact(s)

(C) Pernicious indeterminism

[1] For general reviews of the hole argument, see (Earman, 1989; Stachel, 2014; Norton, 2019; Pooley, 2021). For pre-history of the argument, see (Weatherall, 2020), which also gives some insight into how the trivial fact (that there are distinct but isomorphic models) could have been confused for a substantive fact about the existence of hole isomorphisms.

The upshot, of course, is supposed to weaken the attraction of substantivalism.

Many previous discussions of the hole argument have focused on clarifying what substantivalism means, or on what it means for a theory to be deterministic. We focus instead on pinpointing the claim made in (B), and on checking whether it is true. It seems that there are two mathematical claims that might be relevant here. The first claim is that there are distinct but isomorphic models. But that fact is not strong enough to support the rest of the argument. The second claim is that there are isomorphisms that only move elements inside a hole. But that claim — as we show — is false. In either case, the hole argument fails.

The hole argument is set in the framework of local spacetime theories. The models of such theories have the form $(M, O_1, \ldots, O_n)$, where $M$ is a differentiable manifold and the $O_i$ are tensor fields on $M$. However, saying that makes it immediately sound like the hole argument depends on some deep mathematical subtleties that a typical philosopher would not understand. We do not believe that to be the case, and we will give several analogies for what is going on with the mathematics in the hole argument. The first analogy is to the kind of simple "theory" that one encounters in a logic course. Such a theory consists of a language $\Sigma$ with various symbols, and some axioms in that language. For simplicity, consider the case where $\Sigma$ consists of a single predicate symbol $P$. Then a $\Sigma$-structure consists of a set $M$ and some subset $P^M$ of $M$, the combination of which can be written as $(M, P^M)$.

One important difference between a local spacetime $(M, O_1, \ldots, O_n)$ and a logical model $(M, P^M)$ is that in the former $M$ is a differentiable manifold, whereas in the latter $M$ is a bare set, with no structure relating its elements. (For more details on the theory of differentiable manifolds, see (Lee, 2013; Malament, 2012).) Roughly speaking, an $n$-dimensional manifold $M$ consists of a set (which we again call $M$) and a family $\{(U, \chi)\}$ where $U \subseteq M$ and $\chi$ maps $U$ one-to-one into a subset of the $n$-dimensional real numbers. This family of "charts" gives $M$ enough structure, in the first place, to make sense of the notion of a *smooth* map from $M$ into $\mathbb{R}^k$. In particular, a map $f : M \to \mathbb{R}^k$ is smooth at point $p \in M$ just in case for any chart $(U, \chi)$ with $p \in U$, the composite map $f \circ \chi^{-1}$ is an infinitely differentiable function from $\chi[U]$ to $\mathbb{R}^k$ (this latter notion already being defined in the theory of real numbers). More intuitively, each chart $(U, \chi)$ endows a little patch of $M$ with the structure of $\mathbb{R}^n$, and a smooth map $f : M \to \mathbb{R}^k$ is one such that, restricted to each such patch, is infinitely differentiable. Now if $M$ and $N$

are both differentiable manifolds, then we can also make sense of the idea of a smooth map from $M$ to $N$. In particular, a function $f : M \to N$ is said to be smooth just in case for any function $g : N \to \mathbb{R}^k$, the composite $g \circ f$ is smooth.

**Definition.** Let $M$ and $N$ be manifolds. A function $\varphi : M \to N$ is called a *diffeomorphism* just in case $\varphi$ is smooth and has a smooth inverse $\varphi^{-1}$.

It will be important for what follows to be precise about which structures are, and which structures are not, invariant under diffeomorphisms. Like homeomorphisms of topological spaces, diffeomorphisms need not respect metric structure, and hence they can stretch and deform objects. For example, a diffeomorphism can turn a circle into an ellipse, or a hyperbola into a line. However, diffeomorphisms cannot transform smooth (i.e. infinitely differentiable) curves into lines with kinks. For example, there can be no diffeomorphism $\varphi : \mathbb{R}^2 \to \mathbb{R}^2$ that takes the parabola $\{(x, y) : y = x^2\}$ to the bent line $\{(x, y) : y = |x|\}$.

If $M$ is an $n$-dimensional manifold, then for each point $p \in M$, there is an associated $n$-dimensional real vector space $T_p$ called the *tangent space* at $p$. Intuitively, the tangent space at $p$ represents the possible instant states of motion at $p$ in terms of vectors emanating from $p$. Note, however, that $T_p$ is a bare vector space, which means that the notion of the magnitude of a vector $v \in T_p$ is undefined, as is the notion of the angle between two vectors $v, w \in T_p$. Similarly, if $v \in T_p$ while $w \in T_q$ with $p \neq q$, then "$v$ is the same length as $w$" and "$v$ points in the same direction as $w$" are undefined. In order to define notions like these, one needs further structure, such as a metric on the manifold.

**Definition.** If $V$ is an $n$-dimensional vector space, then a *Minkowski inner product* $\eta$ on $V$ is a mapping that takes pairs of vectors and returns real numbers. We assume, as usual, that there is a one-dimensional subspace of $V$ on which $\eta$ is positive definite, and a $n - 1$ dimensional subspace of $V$ on which $\eta$ is negative definite.

**Definition.** Let $M$ be an $n$-dimensional manifold with $n \geq 2$. A *Lorentzian metric* $g$ on $M$ is a smooth assignment $p \mapsto g_p$ of points in $M$ to Minkowski inner products on the corresponding tangent spaces.

**Definition.** We say that $(M, g)$ is a *Lorentzian manifold*, or alternatively a *relativistic spacetime*, if $M$ is an $n$-dimensional manifold and $g$ is a Lorentzian metric on $M$.

**Definition.** Suppose that $(M, g)$ and $(M, g')$ are relativistic spacetimes and let $\varphi : M \to M'$ be a diffeomorphism. Then $\varphi$ is said to be an *isometry* just in case it preserves the metric structure — i.e. $\varphi^* g' = g$ where $\varphi^*$ is the map which "pulls back" the metric from $M'$ to $M$ (see Malament, 2012, p 36).

It is immediate that if two spacetimes $(M, g)$ and $(M', g')$ are isometric, then their underlying manifolds are diffeomorphic. However, as the following example shows, two spacetimes can fail to be isometric even if their underlying manifolds are diffeomorphic.

**Example.** Let $(M, g)$ be Minkowski spacetime in two dimensions where $M = \mathbb{R}^2$ and $g = dt^2 - dx^2$. Now let $(M', g')$ be the $t > 0$ portion of $(M, g)$, that is, let $M' = \{(t, x) \in M : t > 0\}$ and $g' = dt^2 - dx^2$. We find that the manifolds $M$ and $M'$ are diffeomorphic since the bijection $\varphi : M \to M'$ defined by $\varphi(t, x) = (e^t, x)$ is both smooth and has a smooth inverse. But the diffeomorphism $\varphi$ fails to be an isometry because, when pulling back the metric $g'$ from $M'$ to $M$, we find that $\varphi^* g' = (e^{2t})dt^2 - dx^2 \neq dt^2 - dx^2 = g$. That the two spacetimes have diffeomorphic manifolds but fail to be isometric makes good sense given that they represent radically different physical situations: the truncated Minkowski spacetime $(M', g')$ contains "singularities" in the sense that it is geodesically incomplete while Minkowski spacetime $(M, g)$ is, of course, geodesically complete (Wald, 2010, p 149). □

One can also find non-isometric spacetimes that have quite different causal structures and yet have diffeomorphic underlying manifolds. For a simple example, consider the causally well-behaved (globally hyperbolic) Minkowski spacetime and the causally misbehaved (achronal) Gödel spacetime; each has $\mathbb{R}^4$ as its underlying manifold (see Hawking and Ellis, 1973, p 168; Malament, 2012, p 195). Stepping back, we see that while diffeomorphisms preserve all manifold structure, they need not preserve the most fundamental physical features of a spacetime including its geodesic and causal structure. These sorts of considerations suggest that relativistic spacetimes should be represented by Lorentzian manifolds, and two spacetime models are physically equivalent only if they are isometric.

## 3 Invariance: substantive and trivial

Since there are metrics $g$ and $g'$ on $M = \mathbb{R}^4$ such that $(M, g)$ is geodesically complete and $(M, g')$ is singular, it follows that the property of "being singular" is not invariant under diffeomorphism. In fact not much at all is invariant under diffeomorphism; for example, causal structure is not typically preserved by diffeomorphisms, and tensorial quantities are not typically preserved by diffeomorphisms. These facts might come as a surprise, seeing how often one hears the phrase "diffeomorphism invariance", and how often one hears that only diffeomorphism invariant quantities are physically significant. It would be good, then, to give a more precise accounting of what it means to say that a certain kind of structure is *invariant* under a certain kind of morphism.

If there is one key fact to understanding claims of invariance, it is the following:

> *Any meaningful claim of the form "morphisms of class $\Delta$ preserve structure $S$" presupposes a prior standard of cross-model identity for $S$.*

Conversely, the claim "the morphism $\varphi$ preserves $S$" collapses into triviality if $\varphi$ itself was used to generate the standard of cross-model identity, e.g. by pulling $S^N$ back along $\varphi : M \to N$. Unfortunately, it is the latter, trivial kind of invariance that is at play in claims to the effect that "diffeomorphism invariance" has some special role to play in spacetime physics. In order to see that such claims are trivial, we begin with a couple of simple examples.

**Example.** Consider a language $\Sigma$ that contains a single constant symbol $c$, so that a $\Sigma$-structure $M$ includes the assignment of an element $c^M \in M$. If $(M, c^M)$ and $(N, c^N)$ are $\Sigma$-structures, then a function $f : M \to N$ preserves the relevant structure iff $f(c^M) = c^N$. What is crucial here is that we have an independent grasp of the denotation of $c$ in both $M$ and $N$. If we did not, then we might be tempted to say that *any* function $f : M \to N$ preserves the extension of $c$, because $f$ maps $f^{-1}(c^N)$ to $c^N$, and so we could just let $c^M := f^{-1}(c^N)$. To speak that way is to muddle and trivialize the concept of invariance. □

**Example.** For a slightly more sophisticated example, suppose that $V$ is a vector space, and let $L : V \to V$ be a linear isomorphism. It might be tempting to say that $L$ preserves inner products, because if $m$ is an inner product on $V$, then so is the pulled back inner product $L^*m$ given by

$$(L^*m)(x, y) = m(Lx, Ly).$$

But that way of speaking is deeply confused, because "$L$ preserves inner products" only makes sense if $L$ is a mapping whose domain and range are inner product spaces, i.e. if inner products on both spaces have already been identified. In particular, if $m^V$ is an inner product on $V$, and $m^W$ is an inner product on $W$, then $L : V \to W$ preserves inner products just in case

$$m^W(Lx, Ly) = m^V(x, y), \qquad (\forall x, y \in V),$$

which is the same as saying that $m^v = L^*(m^W)$. To press the point further, note that the definition of $L^*m$ above does not even depend on $L$ being a linear map. Indeed, any bijection of $V$ can be used to define a new linear structure on $V$, and $L^*m$ will be linear relative to this new structure, and the resulting inner-product space will be isomorphic to $(V, m)$. But this does not show that inner products are invariant under arbitrary bijections! $\square$

The same lesson applies to the case of diffeomorphisms and tensorial quantities: the phrase "diffeomorphisms always preserve tensorial quantities" is either meaningless or false. First of all, to speak meaningfully about invariance of tensorial quantities presupposes a fixed standard of cross-model identification of tensorial quantities. (It might help here to remember that a covariant 2-tensor is essentially an indexed family of inner products.) For example, if we define $g^M$ to be the metric tensor of $M$ and $g^N$ to be the metric tensor $N$, then it does makes sense to ask whether some diffeomorphism $\varphi : (M, g^M) \to (N, g^N)$ preserves the metric. And for a general diffeomorphism $\varphi$, the answer will be no. In contrast, the fact that an arbitrary diffeomorphism $\varphi : N \to M$ can be used to pull back a tensor field $g^M$ on $M$ to a tensor field $g^N := \varphi^* g^M$ on $N$ is no more significant than the fact an arbitrary bijection $L : V \to W$ can be used to pull back the structure of an inner-product space $(W, m^W)$. In both cases, the claim that the morphism preserves the structure is trivial, because the structure was defined in terms of the morphism.

To further clarify this point, let's consider the general case of two categories $\mathbf{D}$ and $\mathbf{C}$ of mathematical objects, where the objects of $\mathbf{D}$ have "more structure" than the objects of $\mathbf{C}$. This relation between $\mathbf{D}$ and $\mathbf{C}$ can be captured by saying that there is a "forgetful" functor $U : \mathbf{D} \to \mathbf{C}$, i.e. a functor that is *faithful* but not necessarily *full*. That is, if $f$ and $g$ are distinct morphisms between $M$ and $N$ in $\mathbf{D}$, then $U(f)$ and $U(g)$ are distinct morphisms between $U(A)$ and $U(B)$ in $\mathbf{C}$. However, there may be morphisms between $U(A)$ and $U(B)$ that are not of the form $U(f)$ for some morphism $f : A \to B$.

The familiar concrete categories — such as groups, rings, and fields — fit this mold: if $A$ is an object of $\mathbf{D}$ then $U(A)$ is its underlying set, an object of the category **Set** whose objects are sets and whose arrows are functions. If $A$ and $B$ are objects in $\mathbf{D}$, then there will typically be more morphisms between $U(A)$ and $U(B)$ than there are between $A$ and $B$. To say that a function $f : U(A) \to U(B)$ preserves the structure from the category $\mathbf{D}$ is equivalent to saying that there is a morphism $g : A \to B$ such that $U(g) = f$. Note, however, that this condition is *not* equivalent to saying that there is there is some morphism $h$ of $\mathbf{D}$ such that $U(h) = f$, as the following example shows.

**Example.** Let $\mathbf{D}$ be the category of pointed sets, whose objects are of the form $(X, p)$ with $X$ a set and $p \in X$; and where $f$ is a morphism from $(X, p)$ to $(Y, q)$ just in case $f : X \to Y$ is a function such that $f(p) = q$. Let $\mathbf{C}$ be the category of sets and functions, and let $U$ be the forgetful functor that takes a pointed set $(X, p)$ and returns the set $X$.

Now let $X = \{a, b\}$, and let $f : X \to X$ be the function that permutes $a$ and $b$. Then $f$ is the image under $U$ of a morphism $h$ in $\mathbf{D}$; in particular, $f$ itself is a morphism from $(X, a)$ to $(X, b)$. However, although $f$ is a morphism from $U(X, a)$ to $U(X, a)$, there is no morphism $g : (X, a) \to (X, a)$ such that $U(g) = f$. Thus, $f$ does *not* preserve the structure of pointed sets. $\square$

The schematic of a forgetful functor $U : \mathbf{D} \to \mathbf{C}$ can also be applied in the case where the category $\mathbf{C}$ itself has interesting structure. For example, $\mathbf{C}$ might be the category of topological spaces, while $\mathbf{D}$ is the category of metric spaces, and $U$ is the functor that constructs a topology from a metric. Similarly, $\mathbf{D}$ could be the category whose objects are manifolds with Lorentzian metrics (with isometries as morphisms), while $\mathbf{C}$ is the category of manifolds (with smooth maps as morphisms), and $U$ is the functor that forgets the metric.

When there is a forgetful functor $U : \mathbf{D} \to \mathbf{C}$ then an object $A$ of category $\mathbf{D}$ typically has more invariant structure than the corresponding object $U(A)$ of category $\mathbf{C}$. Nonetheless, there is a trivial sense in which the invariants of the category $\mathbf{D}$ might be said to be preserved by the morphisms of $\mathbf{C}$ (and it is this trivial sense that is at play in Earman and Norton's Gauge Theorem). In particular, given any object $A$ of $\mathbf{D}$ and any isomorphism $f : U(A) \to U(A)$ in $\mathbf{C}$, one can pull back the structure of $A$ along $f$ to create another object

$B$ of $\mathbf{D}$ and an isomorphism $h : B \to A$ such that $U(h) = f$. But this construction does *not* show that the structure of objects in $\mathbf{D}$ is invariant under the morphisms of $\mathbf{C}$; if it did, then any structure that can be built on top of sets would be invariant under all set-theoretic functions. In short, this kind of reasoning would utterly trivialize the concept of invariance.

In summary, if some structure $S$ can be identified across models, then there is a substantive sense in which $S$ might or might not be invariant under some class $\Delta$ of morphisms. In contrast, given some structure $S^M$ on a particular model $M$, any reasonable class of morphisms can be used to pull $S^M$ back to other models, thereby trivializing the claim that $S$ is invariant under that class of morphisms. It is only in this trivial sense that tensorial quantities are invariant under diffeomorphisms. In the substantive sense of invariance, tensorial quantities are *not* invariant under arbitrary diffeomorphisms.

In fact, most the physically interesting features of spacetime models vary under diffeomorphisms. For example, none of the following notions is invariant under diffeomorphism — or to put it more accurately, none of these notions is *definable* in terms of manifold structure alone.

- The length of a curve $\gamma$ in $M$. In particular, if $\varphi : (M, g) \to (M', g')$ is a diffeomorphism, then the length of $\gamma$ (measured by $g$) is not necessarily the same as the length of $\varphi[\gamma]$ (measured by $g'$).

- The property of a vector $v \in T_p$ being timelike. In particular, the pushforward vector $\varphi_* v \in T_{\varphi(p)}$ might be spacelike (according to $g'$) even if the vector $v$ is timelike (according to $g$).

- The fact that $q$ is in the causal future of $p$. In particular, if $\leq$ is the causal ordering on $(M, g)$ and $\preceq$ is the causal ordering on $(M', g')$, then it is possible that $p \leq q$ while not $\varphi(p) \preceq \varphi(q)$.

- The property of $M$ being flat. In particular, a manifold $M$ can be equipped with two metrics $g$ and $g'$, such that $(M, g)$ is flat, but $(M, g')$ is not.

Each of these notions can be defined only relative to some further structure (e.g. a metric $g$); and in such a case, the notion is invariant under morphisms that preserve that structure (e.g. isometries). For this reason, there is a strong *prima facie* case for taking a spacetime to be a pair $(M, g)$ where $M$

is a manifold and $g$ is a Lorentzian metric; and in that case, spacetimes are physically equivalent (i.e. isomorphic) only if there is an isometry between them.

## 4 Representing theoretical commitment

What is it to accept, or believe, the General Theory of Relativity? What is it to be committed to a substantivalist interpretation of the General Theory of relativity? These are big questions, and we will not pretend to give them an adequate treatment in this article. However, we would like to suggest that some aspects of one's theoretical commitments can be represented via the structure of a category of models. This point is by no means novel; indeed, it is taken for granted in most discussions of the foundations of physics; and it has been argued explicitly by Weatherall (2016) and Halvorson (2019). Nonetheless, by emphasizing the point, we can avoid some very basic confusions that arise in the hole argument.

Recall that a category $\mathbf{C}$ consists of a set $C_0$ of objects and a set $C_1$ of arrows between the objects. Each arrow $f \in C_1$ has a domain object $d_0 f$ and a codomain object $d_1 f$. As usual, we write $f : A \to B$ to indicate that $d_0 f = A$ and $d_1 f = B$. Moreover, each object $A \in C_0$ has an identity arrow $1_A$, and if $f : A \to B$ and $g : B \to C$ are arrows, then so is the composite $g \circ f : A \to C$. The structure of the category $\mathbf{C}$ is enough to define the notions of *monomorphism*, *epimorphism*, and *isomorphism*, and it is the last of these notions that plays a central role in the hole argument.

Consider, for example, the following two similar, but inequivalent, categories. Let $\mathbf{C}$ be the category that consists of two objects $A, B$, and that has four arrows: $1_A, 1_B$ and a pair $f : A \to B$ and $g : B \to A$ such that $g \circ f = 1_A$ and $f \circ g = 1_B$. Thus, in the category $\mathbf{C}$, the objects $A$ and $B$ are isomorphic. Now let $\mathbf{D}$ be the category that has the same objects as $\mathbf{C}$, but where $1_A$ and $1_B$ are the only arrows of $\mathbf{D}$. In that case, the objects $A$ and $B$ are non-isomorphic in $\mathbf{D}$.

Now imagine two theories $T_C$ and $T_D$, where the first comes with the category $\mathbf{C}$ (where models $A$ and $B$ are isomorphic) and the second comes with the category $\mathbf{D}$ (where models $A$ and $B$ are not isomorphic). We claim, then, that these two theories have different commitments: $T_C$ recognizes only one possibility (but two ways of representing it) whereas $T_D$ recognizes two possibilities.

Some philosophers might think that $T_C$ would be a defective theory because it has numerically distinct but isomorphic models. In contrast, we don't see this feature of $T_C$ as a defect, nor is it avoidable in practice. Consider, for example, the theory $T$ in first-order logic that says "there are exactly two things." Then $T$ has more set-theoretic models than any cardinal number $\kappa$. However, a person who accepts $T$ is not committing to a claim of the form "there are more possibilities than any cardinal number $\kappa$." No, she is committing to no more nor less than the claim "there are exactly two things". How many possibilities are consistent with her theory is not itself part of her theoretical commitment.

Let us return now to the case of interest, viz. local spacetime theories. Suppose that one theorist, Carsten, has commitments represented by the category **Man** of smooth manifolds and diffeomorphisms, while another theorist, Ditte, has commitments represented by the category **Lor** of manifolds with metric and isometries. In that case, if $g$ and $g'$ are non-isometric metrics on a manifold $M$, then Ditte will take $(M, g)$ and $(M, g')$ to represent distinct possibilities, whereas Carsten will take these two models to be different representations of the same possibility. We argued above that Ditte has a better theory than Carsten, because Carsten cannot even speak coherently about things like causal structure and geodesics. In fact, Carsten considers Minkowski spacetime to be physically equivalent to Gödel spacetime, and also to the Friedman-Robertson-Walker cosmological spacetimes. Thus, Carsten cannot speak coherently about whether the universe is expanding, or whether there are closed timelike curves, etc. Once again, diffeomorphism is far too liberal to be taken as a standard of physical equivalence.

We can now apply this lesson to rule out one last possibility for the structure of the models of GR: let $\mathbf{Man}_g$ be the category whose objects are pairs $(M, g)$ with $M$ a manifold and $g$ a Lorentzian metric, but whose morphisms include all the diffeomorphisms. While the models in $\mathbf{Man}_g$ have metric structure, the morphisms of $\mathbf{Man}_g$ disregard this structure. In fact, the category $\mathbf{Man}_g$ is equivalent to the category **Man** of manifolds and smooth maps. Thus, insofar as theoretical commitment is captured by the structure of the category of models, $\mathbf{Man}_g$ has the same fatal flaw as **Man**, viz. its isomorphisms relate situations that are intuitively physically inequivalent.

In summary, a mature physical theory typically comes equipped with a category of models and morphisms between those models. We have argued that the category of models of GR is not **Man**, because the differentiable manifolds do not have enough structure to be called spacetimes. We have also argued that the category of models of GR is not $\mathbf{Man}_g$, because the isomorphisms in this category ignore physically relevant features of the models. The remaining possibility is that category of models of GR is **Lor**, i.e. the category of Lorentzian manifolds and isometries.

We can imagine, however, an objection to our claim that isometry is the standard of physical equivalence for relativistic spacetimes: don't substantivalists and relationalists employ different standards of physical equivalence? In particular, it is often suggested that substantivalists deny Leibniz equivalence, which suggests that their standard of physical equivalence is more conservative than relationalists' standard of physical equivalence. But how exactly would a substantivalist define an isomorphism between spacetimes? One possibility is that the substantivalist theory includes constant symbols for picking out spacetime points, and so a function $\varphi : X \to Y$ is a substantivalist-isomorphism only if $\varphi(c^X) = c^Y$ for each constant symbol $c$. But that can hardly be the intention of the substantivalist, because in that case he would be committed to Minkowski spacetime having no symmetries. Thus, we see no reason why a substantivalist would adopt a criterion of physical equivalence that is stricter than isometry.

To summarize, the truth of the second premise of the hole argument — i.e. "there are hole isomorphisms" — depends on how one defines "isomorphism", i.e. on which category one considers to be the category of models of GR. We have argued that bare manifolds are far too unstructured to count as spacetimes; and diffeomorphisms ignore far too much structure to count as isomorphisms. While we leave open the possibility that there may be an even better alternative, for the remainder of this paper, we assume that spacetimes are Lorentzian manifolds, and isomorphisms are metric-preserving maps, i.e. isometries.

## 5 The non-existence of hole isomorphisms

We return to the original question of this paper: what mathematical claim is supposed to serve as the second premise of the hole argument? To state it abstractly, the claim is:

($\mathrm{B}_1$) There are relativistic spacetimes $X$ and $Y$, a proper open subset $O$ of $X$, and an isomorphism $\varphi : X \to Y$ that changes things in $O$ but not outside $O$.

(Here we use $X$ instead of $(M, g)$ to remain neutral, for the time being, about the category in which $X$ is considered to be an object.) We first point out a problem with the statement of $\mathrm{B}_1$: how are we to formalize the phrase "changes things"? To see this point more clearly, consider an arbitrary function $\varphi : X \to Y$ between mathematical objects $X$ and $Y$, and imagine being asked to write "$\varphi$ changes things" in mathematical notation. Normally we would say that "$\varphi$ changes things" means that there is an $x \in X$ such that $\varphi(x) \neq x$. But if $X \neq Y$ then the equality relations on $X$ and $Y$ cannot be used to compare $\varphi(x)$ with $x$, and so it is unclear what "$\varphi$ changes things" could mean.

At this point, we have two options. The first option is to introduce an auxiliary morphism $\psi : X \to Y$ to serve as the standard of comparison between elements of $X$ and elements $Y$.[2] Then, "$\varphi$ changes things" could be cashed out as $\varphi \neq \psi$, and $\mathrm{B}_1$ could be re-expressed as:

($\mathrm{B}_2$) There are relativistic spacetimes $X$ and $Y$, a proper open subset $O$ of $X$, and isomorphisms $\psi : X \to Y$ and $\varphi : X \to Y$ such that $\varphi|_{X \setminus O} = \psi|_{X \setminus O}$ but $\varphi|_O \neq \psi|_O$.

The idea behind $\mathrm{B}_2$ is that $\varphi|_{X \setminus O} = \psi|_{X \setminus O}$ expresses "$\varphi$ does not change things outside $O$", while $\varphi|_O \neq \psi|_O$ expresses "$\varphi$ changes things inside $O$".

The second option is to restrict to the special case $Y = X$, where the identity morphism $1_X : X \to X$ can serve as the standard of comparison.[3] In that case, $\mathrm{B}_2$ simplifies to:

($\mathrm{B}_3$) There is a relativistic spacetime $X$, a proper open subset $O$ of $X$, and an isomorphism $\varphi : X \to X$ such that $\varphi|_{X \setminus O} = 1_{X \setminus O}$ but $\varphi|_O \neq 1_O$.

We will soon prove (Theorem 1) that $\mathrm{B}_2$ is false, from which it follows that $\mathrm{B}_3$ is false (Corollary 3). However, we first compare $\mathrm{B}_3$ with the central mathematical claim of Earman and Norton's paper:

**Gauge Theorem** (General Covariance)**.** *If $(M, g)$ is a model of a local spacetime theory and $\varphi$ is a diffeomorphism from $M$ onto $M$, then the carried along tuple $(M, \varphi^* g)$ is also a model of the theory.*

---

[2]The function $\psi : X \to Y$ can be thought of as specifying a "counterpart" relationship (see Butterfield, 1989).

[3]In the hole argument, the relevant spacetime models are of the form $X = (M, \varphi^* g)$ and $Y = (M, g)$. In this case, one might propose the map $1_M : X \to Y$ as the default for "does not change things". However, $1_M$ is a morphism in the category **Lor** only if $g = (1_M)^* g = \varphi^* g$, in which case $X = Y$. If $X \neq Y$, then $1_M : X \to Y$ is not even a physical equivalence, and so it is not a good standard for "does not change things."

The idea here is that $\varphi$ establishes an isomorphism between $(M, \varphi^* g)$ and $(M, g)$; and since the latter is a model of the theory, so is the former. Is this morphism $\varphi$ the sought for hole isomorphism that establishes the truth of $\mathrm{B}_3$?

We claim that the morphism $\varphi$ constructed in the proof of the Gauge Theorem is a hole isomorphism only if $\varphi$ is an isometry. We argue by cases, depending on which category $\mathbf{C}$ the morphism $\varphi$ is supposed to be an isomorphism in: (1) the category $\mathbf{Man}$ of manifolds and smooth maps, (2) the category $\mathbf{Man}_g$ of manifold-metric pairs with smooth maps, or (3) the category $\mathbf{Lor}$ of manifold-metric pairs with isometries.

Case 1: suppose that "$\varphi$ is a hole isomorphism" means that $\varphi$ is an isomorphism in $\mathbf{Man}$. That claim is true, but it is simply not relevant: no reasonable person thinks that a diffeomorphism $\varphi : X \to Y$ establishes that $X$ and $Y$ are physically equivalent. Similarly, no reasonable person thinks that a diffeomorphism $\varphi : X \to X$ shows that there are nomologically possible changes of the situation represented by $X$. But in any case, there is no good reason to think that Earman and Norton intended the Gauge Theorem to show the existence of an isomorphism in $\mathbf{Man}$, because they used the fact that there is an isomorphism in $\mathbf{Man}$ to establish the existence of an isomorphism (but in what category?) between $(M, \varphi^* g)$ and $(M, g)$. At the very least, the notation of the Gauge Theorem suggests that $(M, g)$ is intended to have more structure than just that of a differentiable manifold.

Case 2: suppose that "$\varphi$ is a hole isomorphism" means that $\varphi$ is a morphism in $\mathbf{Man}_g$. We already urged against taking $\mathbf{Man}_g$ as the category of spacetime models, since it is unclear whether the intention would be to represent spacetime as having metric structure or not. But in any case, if $\varphi$ is a hole isomorphism, then $\varphi$ has the same domain and range:

$$(M, \varphi^* g) \;=\; d_0\varphi \;=\; d_1\varphi \;=\; (M, g).$$

But then $\varphi^* g = g$, which means that $\varphi$ is an isometry. In this case, the Gauge Theorem establishes the existence of a hole isomorphism $\varphi$ only if $\varphi : (M, g) \to (M, g)$ is an isometry.

Case 3: suppose that "$\varphi$ is a hole isomorphism" means that $\varphi$ is a morphism in $\mathbf{Lor}$. Once again, $\varphi$ is a hole isomorphism only if $\varphi$ has the same domain and range, and this implies (as in the previous paragraph) that $\varphi$ is an isometry. Therefore, the Gauge Theorem establishes the existence of a hole isomorphism $\varphi$ only if $\varphi : (M, g) \to (M, g)$ is an isometry.

We will soon demonstrate that $\varphi$ could not be an isometry, because there are no isometries that move things around inside a hole, but not outside of it. But let us first pause to develop our intuition for why the Gauge Theorem does not guarantee their existence. Consider the following example.[4]

**Example.** Let $M$ consist of the following three points of the Euclidean plane $a = (0,0), b = (1,0), c = (0,2)$, and let $d$ be the Euclidean metric. Then the metric space $(M,d)$ is rigid in the sense that its only automorphism is $1_M$. (Thus, there most certainly cannot be a hole automorphism $\varphi : (M,d) \to (M,d)$, since there are no non-trivial automorphisms of $(M,d)$.) In particular, there is no isometry of $(M,d)$ that permutes $a$ and $b$. However, if $\varphi : M \to M$ is the map that permutes $a$ and $b$, then we may define a new metric $(\varphi^*d)(x,y) = d(\varphi(x),\varphi(y))$, and then $\varphi$ is an isometry from $(M,\varphi^*d)$ to $(M,d)$. Nonetheless, as we saw above, the automorphism group of $(M,d)$ is trivial. Thus, the fact that there is an isometry from $(M,\varphi^*d)$ to $(M,d)$ does not show that there are any non-trivial automorphisms of $(M,d)$. In fact, we would be within our rights to call the isometry $\varphi : (M,\varphi^*d) \to (M,d)$ trivial, since both $(M,\varphi^*d)$ and $(M,d)$ are rigid, and $\varphi$ is the only isomorphism between them. It would be very strange to say that $\varphi$ "changes things" in $(M,\varphi^*d)$, when that structure has no non-trivial automorphisms. $\square$

The Gauge Theorem does not yet settle the question of whether hole isomorphisms exist, because it does not show that the map $\varphi : (M,g) \to (M,g)$ is an isometry. The mathematical fact needed for the hole argument to go through is the following:

**Conjecture.** *There is a relativistic spacetime $(M,g)$, a proper open subset $O$ of $M$, and an isometry $\varphi : (M,g) \to (M,g)$ that changes things in $O$ but not outside of $O$.*

If this conjecture were true, it would indeed have profound consequences for our understanding of GR. It would entail that everything that happens outside, and in particular in the past, of a tiny region is not sufficient to determine what happens inside that region. If such were true, then GR would display a pernicious form of indeterminism — regardless of whether it was given a substantivalist or relationalist interpretation.

[4] Technically this example falls outside the bounds of the Gauge Theorem, which applies to local spacetime theories. However, a similar example can be constructed by choosing a spacetime $(M,g)$ that has a trivial isometry group. It is easy to find such spacetimes (see Mounoud, 2015).

However, this claim is provably false. Before proving that it is false, we should pause to clarify the sense in which the existence of this kind of hole isomorphism would imply indeterminism. Intuitively, if things in $O$ can be moved around without moving anything in the past of $O$, then there are two spacetimes $(M, g)$ and $(M, g')$ that agree on an initial segment, but that disagree at some later point of time. The existence of such spacetimes would show that GR violates the Montague-Lewis-Earman (MLE) criterion for deterministic theories (see Montague, 1974; Lewis, 1983; Earman, 1986): if possible worlds $W$ and $W'$ agree on some initial segment, then $W = W'$.

There is, however, a problem with the MLE definition, in that it relies on an unclear notion: "two worlds are the same". Does "two words are the same" mean that these worlds are equal *qua* set-theoretic structures, or does it mean that these worlds are isomorphic in some other sense? The former proposal is wildly implausible, because it would entail that *every* theory is indeterministic. Indeed, given a theory $T$ and any model $W$ of $T$, let $W'$ be the model that replaces the spatial points $\Sigma_t$ at time $t$ with primed versions (or any other set $\Sigma'_t \neq \Sigma_t$ of the same cardinality as $\Sigma_t$). Then $W$ and $W'$ agree at all times before $t$, but $W \neq W'$. Therefore, $T$ is indeterministic. (We repeat this argument more formally in the next section.)

We propose to make the MLE definition more precise by replacing the unclear notion of "two worlds are the same" with the unambiguous notion of equality of isomorphisms between models: if a theory is deterministic, then for any two models of that theory, if there is an isomorphism between initial segments of those models, then that isomorphism extends uniquely to the entire models.[5] Here we give a slightly weaker condition which does not guarantee the existence of an extended isomorphism, but does guarantee uniqueness.

**Definition.** We say that theory $T$ has *Property R* just in case for any two models $W$ and $W'$ of $T$, and initial segment $U \subseteq W$, if $f : W \to W'$ and $g : W \to W'$ are isomorphisms such that $f|_U = g|_U$, then $f = g$.

We will not give a general definition of an "initial segment" of a model,

[5]Our Property R is similar to Definition 3 of Belot (1995), which he rejects as inadequate. Since our Property R invokes a comparison morphism $\psi$, it would appear to be based on the notion of "counterparts", whereas Belot suggests — as do Melia (1999) and Teitel (2019) — that no such definition of determinism can capture the full (i.e. haecceitistic) concept. In Section 6 we argue that the kind of determinism they are seeking is a *fata morgana*.

which presupposes that models are equipped with some kind of dynamical structure (but see the next section). For globally hyperbolic spacetimes, the case of primary interest here, an initial segment can be taken to be a Cauchy surface, or the causal past of a Cauchy surface. In any case, the idea behind Property R ("rigidity") is that isomorphisms between spacetime models cannot agree in the past, but disagree in the future.

We will now show that General Relativity (if its models are Lorentzian manifolds) has Property R; and hence, GR admits no hole isometries. We prove, in fact, something stronger: if two isometries agree on any open set, no matter how small, then they agree everywhere.[6]

**Theorem 1.** *Let $(M, g)$ and $(M', g')$ be relativistic spacetimes. If $\varphi$ and $\psi$ are isometries from $(M, g)$ to $(M', g')$ such that $\varphi|_O = \psi|_O$ for some nonempty open subset $O$ of $M$, then $\varphi = \psi$.*

*Proof.* Suppose that $\varphi$ and $\psi$ are isometries and that $\varphi|_O = \psi|_O$ where $O \subseteq M$ is nonempty and open. Consider an arbitrary vector $\xi^a$ at any point $p \in O$. Let $\alpha : V \to \mathbb{R}$ be any smooth map where $V = \varphi[O] = \psi[O]$. Because $\varphi|_O = \psi|_O$, we find that

$$(\varphi_*(\xi^a))(\alpha) = \xi^a(\alpha \circ \varphi) = \xi^a(\alpha \circ \psi) = (\psi_*(\xi^a))(\alpha),$$

where $\varphi_*$ and $\psi_*$ are push forward maps at $p$ (see Malament, 2012). Thus $\varphi_*(\xi^a) = \psi_*(\xi^a)$ for all vectors $\xi^a$ at $p$. Let $\{\xi^a_1, \ldots, \xi^a_4\}$ be an orthonormal tetrad at the point $p$. It follows that $\{\varphi_*(\xi^a_1), \ldots, \varphi_*(\xi^a_4)\} = \{\psi_*(\xi^a_1), \ldots, \psi_*(\xi^a_4)\}$ is an orthonormal tetrad at the point $\varphi(p) = \psi(p)$. From Geroch (1969) we have: If $(M, g)$ and $(M, g')$ are relativistic spacetimes and $\{\xi^a_1, \ldots, \xi^a_4\}$ and $\{\eta^a_1, ..., \eta^a_4\}$ are orthonormal tetrads at points $p \in M$ and $q \in M'$ respectively, then there is at most one isometry $\theta : (M, g) \to (M', g')$ such that $\theta(p) = q$ and $\{\theta_*(\xi^a_1), \ldots, \theta_*(\xi^a_4)\} = \{\eta^a_1, \ldots, \eta^a_4\}$. So there is at most one isometry $\theta : (M, g) \to (M', g')$ such that $\theta(p) = \varphi(p) = \psi(p)$ and

$$\{\theta_*(\xi^a_1), \ldots, \theta_*(\xi^a_4)\} = \{\varphi_*(\xi^a_1), \ldots, \varphi_*(\xi^a_4)\} = \{\psi_*(\xi^a_1), \ldots, \psi_*(\xi^a_4)\}.$$

Since $\varphi$ and $\psi$ are both isometries of this kind, it follows that $\varphi = \psi$. $\square$

[6] The results of this section hold for all Lorentzian manifolds, and not just those satisfying stronger causality conditions, such as global hyperbolicity. However, if a spacetime is *not* globally hyperbolic, then the notion of an "initial segment" may not even be applicable; and so our results fall short of establishing that GR is a deterministic theory (cw. Smeenk and Wüthrich, 2021).

If we take $M' = M$, $g' = g$, and $\psi = 1_M$, then the previous result yields:

**Corollary 2.** *Let $(M, g)$ be a relativistic spacetime. If $\varphi : (M, g) \to (M, g)$ is an isometry that is the identity on some nonempty open subset $O$ of $M$, then $\varphi = 1_M$.*

Since the complement of a hole $O$ in $M$ contains a nonempty open subset of $M$, the previous result entails that there are no hole isomorphisms.

**Corollary 3** (Non-existence of hole isomorphisms)**.** *Let $(M, g)$ be a relativistic spacetime, and let $O$ be a subset of $M$ such that $M \backslash O$ has non-empty interior. If $\varphi : (M, g) \to (M, g)$ is an isometry that is the identity outside of $O$, then $\varphi$ is also the identity inside $O$.*

## 6 Indeterminism: substantive and trivial

We can anticipate what some philosophers will say in response to the results of the previous section. They will say that this kind of result does not really bear upon the hole argument, since that argument is not about the qualitative determinism of GR, but about whether it is fully deterministic, i.e. in a haecceitistic sense. For example, according to Teitel (2019), laws $L$ are *fully deterministic* just in case:

> For all metaphysical possibilities $W$ and $W'$ where $L$ is true, if there is a time $t$ at both $W$ and $W'$ such that $t$ has the same intrinsic properties at both $W$ and $W'$, then $W$ and $W'$ agree on the truth value of every proposition.

If the worlds $W$ and $W'$ agree only on the qualitative propositions — those that do not mention individuals — then the laws are merely *qualitatively deterministic*.[7]

This kind of distinction between full and qualitative determinism is, unfortunately, too vague to be of much use for testing whether theories are deterministic. What does it mean, for example, to say that a proposition mentions individuals? Suppose that there is a property $\phi$ that is instantiated by exactly one individual. Then does the proposition $\exists x \phi(x)$ mention

[7]Teitel's "fully deterministic" and "qualitatively deterministic" correspond, respectively, to the notions **Dm1** and **Dm2** from Butterfield (1989), which in turn correspond to **Det1** and **Det2** from Pooley (2021).

that individual? Or does a proposition have to contain a proper name in order for it to mention individuals? In that case, what are we to do about the fact that most theories in physics do not have such names? Should we say that, for these theories, qualitative determinism automatically implies full determinism? Or should we say that these theories simply cannot be fully deterministic?

Getting caught up on the question of which propositions are qualitative is unlikely to help us to understand GR better, or to understand how GR bears on the substantivalism-relationalism debate. The results of the previous section decisively show that the hole argument does not work for GR, because GR — whether interpreted substantivally or relationally — has dynamically rigid models. Why is it, then, that there has been, and will surely continue to be, a feeling that there is some remaining open question about whether GR is fully deterministic? Our conjecture is that the worry here arises from the fact that GR, just like any other theory of contemporary mathematical physics, allows its user a degree of representational freedom, and consequently displays a kind of *trivial semantic indeterminism*: how things are represented at one time does not constrain how things must be represented at later times.

To see this point more clearly, we consider a few examples from (many-sorted) first order logic (see Halvorson, 2019).[8]

**Definition.** Let $I$ be a linear order. We say that $T$ is a *dynamical theory* just in case $T$ has sort symbols $\{\sigma_i : i \in I\}$, and a set (possibly empty) of relation or function symbols $\delta_{ij}$ of sort $\sigma_i \to \sigma_j$ specifying transitions from time $i$ to time $j$.

Of course, this definition of a dynamical theory could be generalized in various ways, e.g. by allowing $I$ to be a partial order, or by allowing the $\delta_{ij}$ to be relation symbols instead of function symbols. However, the simple type of dynamical theory specified here will be sufficient for our argument.

If $T$ is a dynamical theory, then in any model $M$ of $T$, the set $M_i = M(\sigma_i)$ will be understood as the individuals, or spatial points, that exist at time $i$. We can now generalize the definition of a theory with Property R to any dynamical theory.

**Definition.** Let $I$ be a linear order. We say that $U \subseteq I$ is an *initial segment* if $U$ is nonempty and for any $j \in U$, if $i \in I$ and $i \leq j$, then $i \in U$.

[8]Recall that every theory with finitely many sorts can be converted to a single-sorted theory (see Barrett and Halvorson, 2017); so the examples we use here could be rewritten as single-sorted theories.

**Definition.** Let $T$ be a dynamical theory. We say that $T$ has *Property R* just in case for any two models $M$ and $N$ of $T$, and any two isomorphisms $f, g : M \to N$, if $f_i = g_i$ for all $i$ in some initial segment $U$, then $f = g$.

Let us compare Property R with the Montague-Lewis-Earman definition of a deterministic theory. First of all, if $M$ and $N$ are complete histories, i.e. models of $T$, then we cash out "$M$ and $N$ agree on intrinsic properties throughout some initial segment $U$" as: for all $i \in U$, the domains $M_i$ and $N_i$ are identical, and for any relation $R$ between individuals at time $i$, we have $R^M = R^N$. (Here we can cash out the idea that $R$ is a relation between individuals at time $i$ by specifying that $R$ is of sort $\sigma_i \times \cdots \times \sigma_i$.) We will write $M|_U = N|_U$ to express that $M$ and $N$ agree on the initial segment $U$, which case MLE determinism can be defined as follows:

**Definition.** Let $T$ be a dynamical theory. We say that $T$ is *deterministic in the sense of Montague-Lewis-Earman* just in case for any two models $M$ and $N$ of $T$, if there is an initial segment $U$ such that $M|_U = N|_U$, then $M = N$.

But in this case, we have the following result:

**Proposition.** *No dynamical theory is deterministic in the sense of Montague-Lewis-Earman.*

*Proof.* Let $T$ be a dynamical theory, let $M$ be a model of $T$, fix a non-trivial initial segment $U \subseteq I$, and choose $i \notin U$. Now let $N_i$ be a set that has the same cardinality as $M_i$, but that is not identical to $M_i$, and let $N$ be the model of $T$ that is just like $M$ except where $M_i$ is replaced with $N_i$. Then $M|_U = N|_U$ but $M \neq N$, showing that $T$ is not MLE deterministic. $\square$

The construction in the foregoing proof is not very interesting: it just uses the fact that for any set $M_i$, there is an isomorphic but non-identical set $N_i$. However, it is not difficult to construct more interesting examples of theories that have dynamically rigid models (i.e. that satisfy Property R) but that fail to be MLE deterministic.

**Example.** Let $\Sigma = \{\sigma_0, \sigma_1, \delta\}$, where $\sigma_0$ and $\sigma_1$ are sort symbols, and $\delta$ is a function symbol of sort $\sigma_0 \to \sigma_1$. Thus, a $\Sigma$ structure $M$ consists of two sets $M_0$ and $M_1$ and a function $\delta^M : M_0 \to M_1$. Now let $T$ be the theory in signature $\Sigma$ that says that:

1. There are exactly two things of sort $\sigma_0$ and exactly two things of sort $\sigma_1$, and

2. $\delta$ is a bijection.

Intuitively, $T$ says that there are two times ($t_0$ and $t_1$), and that at each time, there are two spatial points, such that each point at $t_0$ is connected to a unique point at $t_1$. So, if $M$ is a model of $T$, then we can think of a pair $\langle a, \delta^M(a)\rangle$ with as the unique "geodesic" passing through $a \in M_0$. Furthermore, geodesics are invariant under isomorphism in the sense that if $h : M \to N$ is an isomorphism of models, then

$$\langle h(a), h(\delta^M(a))\rangle \;=\; \langle h(a), \delta^N(h(a))\rangle .$$

There is a precise sense, then, in which the models of $T$ have determinate dynamical structure. The models of GR have similarly determinate dynamical structure: for each globally hyperbolic spacetime $(M, g)$, and for each $p \in M$, there is a unique inextendible geodesic $\gamma$ passing through $p$; and if $h : (M, g) \to (M', g')$ is an isometry, then $h[\gamma]$ is the unique inextendible geodesic passing through $h(p)$. $\square$

To make the previous example a bit more interesting, we could add the following:

**Example.** Suppose that $T^+$ is the extension of $T$ that includes a predicate symbol $P$, and axioms that say that one and only one object is $P$ at each time, and that $P$ is preserved by $\delta$. That is,

$$\forall x(P(x) \to P(\delta(x))).$$

(Technically, there should be a predicate $P_0$ of sort $\sigma_0$, and another predicate $P_1$ of sort $\sigma_1$. But we omit that complication.) Then $T^+$ is deterministic in the following sense: in any fixed model $M$ of $T^+$, if $a$ is $P$ at time 0, then $\delta(a)$ is $P$ at time 1. $\square$

This theory $T^+$ is an interesting test case for one's intuitions about determinism. First of all, $T^+$ is indeterministic in the sense of Montague-Lewis-Earman: it has models $M$ and $N$ that agree at the initial time $t_0$ but then disagree at time $t_1$. For example, let:

$$\begin{array}{ll} M_0 = \{a, b\} & N_0 = \{a, b\} \\ M_1 = \{c, d\} & N_1 = \{c, d\} \\ M_0(P) = \{a\} & N_0(P) = \{a\} \\ M_1(P) = \{c\} & N_1(P) = \{d\} \\ \delta^M(a) = c & \delta^N(a) = d \end{array}$$

Thus, $M$ and $N$ agree on all facts at time $t_0$, but $M$ and $N$ disagree on which object has property $P$ at time $t_1$.

But it would be strange to call the theory $T^+$ "indeterministic", especially since it is a categorical theory, i.e. it has a unique model up to isomorphism. In fact, for any two models $M$ and $N$ of $T^+$, there is a *unique* isomorphism $f : M \to N$. This fact leads us to think that the models $M$ and $N$ described above only apparently disagree about the facts at time $t_1$. The object $d \in N_1$ is the isomorphic image of $c \in M_1$, which suggests to us that "$c$" (in $M_1$) names the same object as "$d$" (in $N_1$). So perhaps we need not take the element $c$ in the model $M$ to be representing the same physical object as the element $c$ in the model $N$.

The question before us is whether a theory's failing the Montague-Lewis-Earman criterion is a sign that that theory is not as deterministic as a theory could be. We think not. As we have seen, a theory can fail to be MLE deterministic simply because the language of set-theory allows the same situation to be described in different ways. (That is precisely what set theory is good at doing: constructing new sets.) In particular, the theories $T$ and $T^+$ display this trivial semantic indeterminism even though both of them are deterministic in the following precise sense:

**Proposition.** *The theories $T$ and $T^+$ have Property R.*

*Proof.* We treat the case of $T$; an analogous argument works for $T^+$. Suppose that $M$ and $N$ are models of $T$, and that $f, g : M \to N$ are isomorphisms such that $f_0 = g_0$. By the definition of homomorphisms of models (Halvorson, 2019, p 180), the following diagram commutes.

$$\begin{tikzcd}
M_1 \arrow[r, shift left, "f_1"] \arrow[r, shift right, "g_1"'] & N_1 \\
M_0 \arrow[u, "\delta^M"] \arrow[r, shift left, "f_0"] \arrow[r, shift right, "g_0"'] & N_0 \arrow[u, "\delta^N"']
\end{tikzcd}$$

Since $f_0 = g_0$, it follows that $\delta^N \circ f_0 = \delta^N \circ g_0$, and hence $f_1 \circ \delta^M = g_1 \circ \delta^M$. Since $\delta^M$ is an epimorphism, $f_1 = g_1$. Hence, if $M$ and $N$ are models of $T$, and $f, g : M \to N$ are isomorphisms, then $f_0 = g_0$ only if $f = g$. It follows that $T$ has Property R. $\square$

Of course, there are dynamical theories that lack Property R; and these theories, we claim, are genuinely indeterministic. For example, let $T'$ be the theory that is just like $T$, but that drops the assumption that there is a dynamical map $\delta$ connecting objects at time 0 to objects at time 1. To be precise, let $\Sigma' = \{\sigma_0, \sigma_1\}$, and let $T'$ be the theory in $\Sigma'$ that says that there are two objects of each sort. Unlike $T$, this theory $T'$ is genuinely indeterministic in the following sense: there are models $M$ and $N$ of $T'$, and isomorphisms $f, g : M \to N$ such that $f_0 = g_0$ but $f \neq g$. Indeed, any bijection $f_0 : M_0 \to N_0$ can be combined with any bijection $f_1 : M_1 \to N_1$ to give an isomorphism $f : M \to N$. This shows that models of $T'$ do *not* have determinate dynamical structure that connects earlier states of affairs to later states of affairs.

Note that the difference between the deterministic $T$ and the indeterministic $T'$ has nothing to do with the latter being substantivalist. The two theories make exactly the same existence claims, and the deterministic theory $T$ actually posits more structure than the indeterministic theory $T'$. (In general, a theory posits more structure if its category of models has fewer morphisms.) The lesson here is that indeterminism does not arise from the claim that spacetime is a substance, but from the claim that spacetime *lacks* objective dynamical structure. If the hole argument undercuts any position regarding the metaphysics of spacetime, it is the sort of metric anti-realism that was championed by Reichenbach and Grünbaum (see Putnam, 1963).

On the reading that we have urged, GR is a metrically realist theory, i.e. its models have determinate metric structure. And in that case, GR is as deterministic as a theory can be.[9] Why is it, then, that some philosophers still are not convinced? What exactly is it that they want, but that our Property R does not provide? We have thought long and hard about this question, and have come to the conclusion that our opponent is not thinking about the theory GR, but about GR+$\mathrm{ZF}_m$, where the latter is ZF set theory read in the "material mode" as a theory about concrete possibilia. Let us explain.

Consider a simple theory T that says "there is exactly one thing," and consider two models: $\{a\}$ and $\{b\}$. Now there are two very different ways to think about these models:

> (material mode) $\mathrm{ZF}_m$ implies that $a \neq b$, and so $\{a\} \neq \{b\}$.
>
> (formal mode) $a$ and $b$ are names, and either one could be used to name the unique thing that exists, according to T. The latter theory does not say that $a \neq b$, or that $\{a\} \neq \{b\}$. In fact, T does not make any assertions about the identity or distinctness of models.

Thus, for a person who accepts only T and its ideology, the question "are $\{a\}$ and $\{b\}$ the same world?" cannot even be formulated. In a similar way, for a person who accepts only GR and its ideology, the statement "$p$ can have different metric properties in different models" cannot even be formulated. Clearly, then, GR+$\mathrm{ZF}_m$ and GR are different theories, with different expressive capacities; and GR+$\mathrm{ZF}_m$ might fail to be deterministic even when GR itself is deterministic.

Why is it then that our opponent wants to talk about GR+$\mathrm{ZF}_m$ instead of just GR? Or is it the spacetime substantivalist who is committed to GR+$\mathrm{ZF}_m$? We argued earlier that spacetime substantivalists do *not* need to enrich GR with names for spacetime points. For similar reasons, spacetime substantivalism can get along just fine without the tendentious theory of possibilia supplied by $\mathrm{ZF}_m$. The theory GR already quantifies over spacetime points, and that should be enough for substantivalists.

To be fair, it is not just in discussions of the hole argument that $\mathrm{ZF}_m$ gets smuggled in the back door. In fact, it has become something of a habit of analytic philosophers to translate the formal-mode claims of set-theoretic

[9] To be clear, our claim here is meant only to apply to GR with globally hyperbolic spacetimes.

model theory into material-mode claims of modal metaphysics. And why do philosophers do this? The reason, we think, is because they worry that if semantic ascent is not halted at the point of set theory, then no positive assertions will ever be made.

We agree about the need to halt semantic ascent, but we would propose halting it one step earlier in the sequence. That is, instead of using the language of $ZF_m$ to express assertions of metaphysical interest, let's use the language of physical theories themselves. And what can be said with the language of GR? As a rough guide, the language of GR allows us to say the sort of things that expert users of GR say about the external world — e.g. "there is an inextendible geodesic of finite length" or "if the mass increases beyond a certain bound, then a singularity will form." If one worries that this account is too vague, then we could, of course, engage in a project of regimenting GR. What that project would show is that the language of GR has words such as "metric" and "geodesic", but no names for spacetime points. So, when a philosopher starts talking about spacetime points having different properties in different possible worlds, then they have already gone beyond the language of GR.

## 7 Metric essentialism

We are not the first ones to suspect that the hole argument can be blocked by metric realism, i.e. the claim that spacetime has metric structure. In fact, to give credit where it is due, we got the idea from reading Maudlin (1988; 1990), where he says:

> Earman and Norton's difficulty arises from asserting that the substantivalist must regard space-time as represented by the bare topological manifold. (Maudlin, 1990, p 545)

> The substantivalist's natural response to the hole dilemma is to insist that spacetime is represented not by the bare manifold but by the manifold plus metric, by the metric space. (Maudlin, 1990, p 546)

We agree with Maudlin that spacetime should be represented by a manifold with metric, but we disagree with his motivation for this claim. Maudlin's motivation for this claim is *metric essentialism*, according to which "spacetime points possess their metric properties essentially." However, metric essentialism is not needed to motivate metric realism; indeed, we believe that metric essentialism yields a worse theory than simple metric realism.

Consider the following simple analogy. Suppose that I have a theory $T$ according to which there are exactly two people, and exactly one of those people has blond hair. According to standard model-theoretic semantics, $T$ has many different models, but all of these models are isomorphic. For example, $T$ has a model where Alice has blond hair and Bob does not; and $T$ has another (isomorphic) model where Bob has blond hair and Alice does not. The theory $T$ has just the right amount of structure to say that a person's hair color is an objective fact, without committing to extraneous claims, such as the claim that hair color is an essential property.

Suppose now that Tim has a theory $T'$ that is just like $T$, except that $T'$ includes the claim that hair color is an essential property. But then if $T'$ has a model where Alice has blond hair, then it cannot have a model where Bob has blond hair; because if Alice has blond hair in one model, then she has blond hair in all models. However, Tim is now in an awkward position: he does not know what the models of his theory are until he determines which person has blond hair. So, to the extent that knowing a theory is knowing what possibilities it permits, Tim does not even know his own theory. In contrast, I know exactly which possibilities my theory permits.

Of course, Tim could double down and complicate his theory: he could say that $T'$ has two collections of models that are, in some sense, mutually inaccessible. It would be perfectly coherent to develop this idea, but one wonders what advantage is to be gained. Why do we need essential properties when regular properties will do the work?

Returning now to the case of local spacetime theories, and specifically to GR: on the most obvious reading, GR makes no claims to the effect that if a point $p$ has a metric profile $\phi$ in one world, then it has profile $\phi$ in all worlds. To make such a claim, GR would either need names for points (to identify them across worlds), or it would need modal operators to say things like $\forall x(\phi(x) \to \Box\phi(x))$. Lacking both of these resources, GR makes no claims about spacetime points possessing their metric properties essentially.[10]

[10]There is, however, a truth in the neighborhood of what Maudlin is saying: isomorphisms preserve a theory's predicates, and ascriptions of metric values are among the predicates of GR. In particular, if $h : (M, g) \to (M', g')$ is an isomorphism, then for any $a \in M$, $(M, g) \vDash \phi(a)$ only if $(M', g') \vDash \phi(h(a))$. But this latter mathematical fact falls

To summarize, Maudlin correctly intuited that the hole argument gains plausibility by ignoring the metric structure of spacetime. We have validated his intuition by showing that if spacetime does have metric structure, then there are no hole isomorphisms. However, the claim that spacetime has metric structure is a good deal weaker, and more plausible, than the claim that spacetime points have their metric properties essentially. We see no need for the latter claim; and indeed, we see it as a superfluous addition to the content of the General Theory of Relativity.

## 8 Conclusion

The hole argument is supposed to show that spacetime substantivalism implies indeterminism. What's more, the notion of indeterminism at play is that of Montague, Lewis, and Earman: there are possible worlds that agree on an initial segment but then later diverge. However, what it means to say that possible worlds are the same, or that they agree on an initial segment, was not specified precisely. In fact, there are distinct notions of isomorphism at play; and the hole argument is seductive only because it equivocates between these different notions of isomorphism.

No reasonable person would adopt a theory where spacetime only has the structure of a differentiable manifold — for in that case, there would not be enough structure to define the most basic spatio-temporal notions. Similarly, no reasonable person would adopt a theory where spacetime has some structure, but where two spacetime models can be "physically equivalent" even while differing with respect to that structure. Thus, the existence of a diffeomorphism between spacetime models fails to establish that those models are physically equivalent.

Our own preferred regimentation of GR is that spacetime models have metric structure, and that physical equivalence of models is established by isometry. In that case, there are no hole isomorphisms, and hence, no mathematical fact that can support the hole argument.

*Acknowledgments:* For helpful feedback on earlier drafts, we thank Thomas Barrett, Brendan Kolb, Chris Smeenk, Jim Weatherall, and two anonymous referees. HH thanks Des Hogan for discussion about metric essentialism.

---

short of establishing that $\phi$ is an essential property of $a$. For one, the relation between $a$ and $h(a)$ is not one of identity; indeed, $h$ is one of possibly many isomorphisms between $(M,g)$ and $(M',g')$, and $a$ cannot be identical to all of its images under isomorphism.

## References


Barrett, Thomas William and Hans Halvorson (2017). "Quine's conjecture on many-sorted logic". In: *Synthese* 194.9, pp. 3563–3582. DOI: 10.1007/s11229-016-1107-z.

Belot, Gordon (1995). "New work for counterpart theorists: Determinism". In: *The British journal for the philosophy of science* 46.2, pp. 185–195.

Butterfield, Jeremy (1989). "The hole truth". In: *The British Journal for the Philosophy of Science* 40.1, pp. 1–28. DOI: 10.1093/bjps/40.1.1.

Earman, John (1986). *A Primer on Determinism*. Springer.

— (1989). *World Enough and Spacetime*. MIT press.

Earman, John and John Norton (1987). "What price spacetime substantivalism? The hole story". In: *The British Journal for the Philosophy of Science* 38.4, pp. 515–525. DOI: 10.1093/bjps/38.4.515.

Geroch, Robert (1969). "Limits of spacetimes". In: *Communications in Mathematical Physics* 13.3, pp. 180–193. DOI: cmp/1103841574.

Halvorson, Hans (2019). *The Logic in Philosophy of Science*. Cambridge University Press. DOI: 10.1017/9781316275603.

Hawking, Stephen and George Ellis (1973). *The Large Scale Structure of Space-time*. Cambridge University Press.

Lee, John M (2013). *Smooth Manifolds*. Springer. DOI: 10.1007/978-1-4419-9982-5.

Lewis, David (1983). "New work for a theory of universals". In: *Australasian Journal of Philosophy* 61.4, pp. 343–377. DOI: 10.1080/00048408312341131.

Malament, David B (2012). *Topics in the Foundations of General Relativity and Newtonian Gravitation Theory*. University of Chicago Press. DOI: 10.7208/chicago/9780226502472.001.0001.

Maudlin, Tim (1988). "The essence of space-time". In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*. Vol. 1988. 2. Philosophy of Science Association, pp. 82–91. DOI: 10.1086/psaprocbienmeetp.1988.2.192873.

— (1990). "Substances and space-time: What Aristotle would have said to Einstein". In: *Studies in History and Philosophy of Science Part A* 21.4, pp. 531–561. DOI: 10.1016/0039-3681(90)90032-4.

Melia, Joseph (1999). "Holes, haecceitism and two conceptions of determinism". In: *The British journal for the philosophy of science* 50.4, pp. 639–664.

Montague, Richard (1974). “Deterministic theories”. In: *Formal Philosophy*. Ed. by Richmond Thomason. Yale University Press, pp. 303–359.

Mounoud, Pierre (2015). “Metrics without isometries are generic”. In: *Monatshefte für Mathematik* 176.4, pp. 603–606. DOI: 10.1007/s00605-014-0614-6.

Norton, John (2019). “The hole argument”. In: *The Stanford Online Encyclopedia of Philosophy*. URL: https://plato.stanford.edu/entries/spacetime-holearg.

Pooley, Oliver (2021). “The hole argument”. In: *The Routledge Companion to Philosophy of Physics*. Ed. by Eleanor Knox and Alastair Wilson. Routledge, pp. 145–159. DOI: 10.4324/9781315623818-15.

Putnam, Hilary (1963). “An examination of Grünbaum’s philosophy of geometry”. In: *Philosophy of Science, The Delaware Seminar*. Vol. 2, pp. 205–255.

Smeenk, Chris and Christian Wüthrich (2021). “Determinism and general relativity”. In: *Philosophy of Science* 88.4, pp. 638–664. DOI: 10.1086/713904.

Stachel, John (2014). “The hole argument and some physical and philosophical implications”. In: *Living Reviews in Relativity* 17.1, pp. 1–66.

Teitel, Trevor (2019). “Holes in spacetime: Some neglected essentials”. In: *The Journal of Philosophy* 116.7, pp. 353–389. DOI: 10.5840/jphil2019116723.

Wald, Robert M (2010). *General Relativity*. University of Chicago Press.

Weatherall, James Owen (2016). “Are Newtonian gravitation and geometrized Newtonian gravitation theoretically equivalent?” In: *Erkenntnis* 81.5, pp. 1073–1091. DOI: 10.1007/s10670-015-9783-5.

— (2018). “Regarding the ‘hole argument’”. In: *The British Journal for the Philosophy of Science* 69.2, pp. 329–350. DOI: 10.1093/bjps/axw012.

— (2020). “Some philosophical prehistory of the (Earman-Norton) hole argument”. In: *Studies in History and Philosophy of Modern Physics* 70, pp. 79–87. DOI: 10.1016/j.shpsb.2020.02.002.