What Hole Argument?

Hans Halvorson and JB Manchak

May 12, 2021

In terms of generating discussion, few articles in the philosophy of physics can parallel Earman and Norton's (1987) article on the "hole argument" in the General Theory of Relativity. In short, by the 1970s, spacetime substantivalism had come into vogue, but Earman and Norton argued that a substantivalist must be committed to a pernicious form of indeterminism. Their argument seems to cleverly exploit the diffeomorphism freedom of GTR, a mathematical subtlety that had tripped up even Einstein himself.

Here we attempt to answer the following question: what is the mathematical fact or facts on which the hole argument is supposed to be based? We identify two mathematical claims that might be relevant. The first of these mathematical claims is trivially true — as pointed out by Weatherall (2018) — and so does not underwrite any metaphysically interesting conclusions. While we agree with Weatherall's point, we suggest that others may have confused the trivial mathematical fact he identifies with another mathematical claim, which, if true, would have profound consequences for the interpretation of GTR. We prove here that this second mathematical claim is false, and we conclude that there is no basis for the hole argument.

1 Uninteresting facts

We begin by restating the trivial mathematical fact that might mistakenly be taken to generate the hole problem. In what follows, if M is a structured set (i.e. an object in a concrete category) then we let $|M|$ denote its underlying set.

Uninteresting Fact 1. *Let M be a structured set, and let $f : |M| \to |M|$ be any bijection such that f is not an automorphism of M. Then one can pull* *back the structure of M along f to define a distinct structured set N, such that $|M| = |N|$ and $f : N \to M$ is an isomorphism.*

As Weatherall (2018) convincingly argued, Fact 1 is uninteresting from the point of view of physics and metaphysics. We will just add a little bit of our own gloss to his points.

The elements of $|M|$ are mathematical objects, and our freedom to permute them is the freedom of the mathematician. The elements of $|M|$ can be used to represent things in the physical world, but in that case, a permutation $f : |M| \to |M|$ has no prima facie significance beyond a change of notation.

More importantly, the existence of the isomorphism $f : N \to M$ supplies no interesting information about the structure of the category of models of the relevant physical theory. By "interesting information" we mean, for example, information about whether the theory's models have non-trivial automorphisms, i.e. bijections that preserve the relevant structure. In particular, the existence of an isomorphism $f : N \to M$ does not show that M itself has any non-trivial automorphisms.

There is a second mathematical fact that is only slightly less trivial than the first. For this second fact, the notion of a permutation of the underlying set is replaced with the notion of a diffeomorphism of the underlying manifold.

Uninteresting Fact 2. *Let M be a smooth manifold, and let g be a metric on M. Given a proper open subset O of M, there is a diffeomorphism $\varphi : M \to M$ that is non-trivial inside O, but trivial outside of O. Hence, φ is an isometry from $(M, \phi^* g)$ to (M, g), where φ^* is the pullback of covariant tensors.*

The commonly used notation (M, g) is the source of much confusion, for it tempts one to think of φ as a non-trivial isometry from M to itself. But M is a bare manifold, not a manifold with metric, and so it does not make sense to talk about isometries of M. Instead, we should be clear that the domain and range of the isometry φ are distinct objects in the category of Lorentzian manifolds. Moreover, there is no more reason to think of this isometry φ as non-trivial than there was to think of the isomorphism $f : N \to M$ from Fact 1 as non-trivial. In particular, the existence of the isometry $\varphi : (M, \phi^* g) \to (M, g)$ does not imply that (M, g) has any non-trivial automorphisms. For example, if (M, g) is a rigid spacetime (i.e. one that has no non-trivial isometries) then $(M, \varphi^* g)$ is also a rigid spacetime, and the map φ is no more physically interesting than the identity automorphism of (M, g).

2 Interesting non-facts

It is a trivial fact that the same mathematical object can be represented in different ways, say as (M, g) and as $(M, \phi^* g)$. So how could that fact generate so much philosophical discussion? We suggest that the trivial mathematical fact was confused for the following:

Interesting Claim. *There is a relativistic spacetime M, a proper open subset O of M, and an isometry $f : M \to M$ that acts non-trivially inside O, but trivially outside O.*

If this claim were true, it would have profound consequences for our understanding of GTR. It would indeed entail that everything that happens outside, and in particular in the past, of O is not sufficient to determine what happens inside O. If such were true, we would be forced to admit that GTR is indeterministic in one very precise sense. Before proving that this claim is false, we should pause to clarify the sense in which the existence of this kind of "hole isometry" would imply indeterminism.

Intuitively, if things in O can be moved around without moving anything in the past of O, then there are two distinct spacetimes M and M' that agree on an initial segment. The existence of such spacetimes would show that GTR 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, namely that it relies on an imprecise account of the identity of worlds (or models, or spacetimes). Indeed, consider again the problems we ran into with the Uninteresting Facts from earlier in this paper. In that case there are two models M and N that are isomorphic. So should we consider M and N to be the same or different models? The MLE definition of determinism provides no guidance on this issue, and so it is all too easy to arrive at contradictory verdicts.

We propose to make the MLE definition more precise by replacing the ambiguous notion of sameness of worlds with the unambiguous notion of equality of isomorphisms between models. The basic idea here should be:

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. 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 M and N of T, and initial segment $U \subseteq M$, if $f : M \to N$ and $g : M \to N$ are isomorphisms such that $f|_U = g|_U$, then $f = g$.

We will not try to give a general definition of an "initial segment" of a model, which presupposes that models are equipped with some kind of dynamical structure. For globally hyperbolic spacetimes, the case of primary interest here, an initial segment can be taken to be the causal past of some Cauchy surface. In any case, the idea behind Property R ("rigidity") is that isomorphisms between M and N cannot agree in the past, but disagree in the future.

We are now prepared to show that relativistic spacetimes have Property R, and hence, that the "interesting claim" is false. That is, there is no relativistic spacetime that permits a non-trivial hole isometry. We prove, in fact, something much stronger: if two isometries agree on any open set, no matter how small, then they agree everywhere.

Proposition. *Let M and N be relativistic spacetimes. If $f : M \to N$ and $g : M \to N$ are isometries such that $f|_U = g|_U$ for some non-empty open subset $U \subseteq M$, then $f = g$.*

Proof. Let M and N be relativistic spacetimes. Let $f : M \to N$ and $g : M \to N$ be isometries such that $f|_U = g|_U$ for some open subset $U \subseteq M$. Consider an arbitrary vector ξ^a at any point $p \in U$. Let $\alpha : V \to \mathbb{R}$ be any smooth map where $V = f[U] = g[U]$. Because $f|_U = g|_U$, we find that $(f_*(\xi^a))(\alpha) = \xi^a(\alpha \circ f) = \xi^a(\alpha \circ g) = (g_*(\xi^a))(\alpha)$ where f_* and g_* are push forward maps at p (see Malament, 2012). Thus $f_*(\xi^a) = g_*(\xi^a)$ for all vectors ξ^a at p. Let $\{\xi^a_1, ..., \xi^a_4\}$ be an orthonormal tetrad at the point p. It follows that $\{f_*(\xi^a_1), ..., f_*(\xi^a_4)\} = \{g_*(\xi^a_1), ..., g_*(\xi^a_4)\}$ is an orthonormal tetrad at the point $f(p) = g(p)$. From Geroch (1969) we have: If M and N are relativistic spacetimes and $\{\xi^a_1, ..., \xi^a_4\}$ and $\{\eta^a_1, ..., \eta^a_4\}$ are orthonormal tetrads at points $p \in M$ and $q \in N$ respectively, then there is at most one isometry $\varphi : M \to N$ such that $\varphi(p) = q$ and $\{\varphi_*(\xi^a_1), \ldots, \varphi_*(\xi^a_4)\} = \{\eta^a_1, ..., \eta^a_4\}$. So there is at most one isometry $\varphi : M \to N$ such that $\varphi(p) = f(p) = g(p)$ and $\{\varphi_*(\xi_1^a), ..., \varphi_*(\xi_4^a)\} = \{f_*(\xi_1^a), \ldots, f_*(\xi_4^a)\} = \{g_*(\xi_1^a), ..., g_*(\xi_4^a)\}$. Since f and g are both isometries of this kind, we know $f = g$. $\square$

From the previous result, we immediately obtain the following by taking $M = N$ and $g = 1_M$.

Corollary. *Let M be a relativistic spacetime. If $f : M \to M$ is an automorphism that is the identity on some non-empty open subset $U \subseteq M$, then f is the identity automorphism.*

In particular, there is no way to isometrically move the contents of a proper open subset O of M without moving something in the causal past of O.

Crucially, our argument depends on the assumption that the relevant notion of isomorphism between spacetimes is isometry. While we could imagine some applications where other notions of sameness might be relevant, we cannot imagine a case where diffeomorphism would be the relevant notion. Diffeomorphisms of spacetimes preserve only their smoothness structure, and can completely alter the most fundamental physical features of a spacetime, including its causal structure. For example, Minkowski spacetime — which is the causally best behaved spacetime imaginable — is diffeomorphic to a spacetime with a singularity. Thus, if one claims that diffeomorphism is a relevant notion of sameness, then one does not believe that singularities are physically real features of spacetimes.

3 Trivial semantic indeterminism

In this section, we give an example of two simple theories that illustrate the difference between genuine indeterminism and "trivial semantic indeterminism." By the latter we mean that any theory can count as indeterministic insofar as we can change our minds about how to use words. For convenience, our examples are formulated in many-sorted logic (see Halvorson, 2019).

Let $\Sigma_1 = \{\sigma_1, \sigma_2, \gamma\}$, where σ_1 and σ_2 are sort symbols, and γ is a function symbol of sort $\sigma_1 \to \sigma_2$. Let T_1 be the theory in signature Σ_1 that says that γ is a bijection. The intended interpretation of T_1 is that σ_1 is the sort of spatial points at one (earlier) time, σ_2 is the sort of spatial points at another (later) time, and $\gamma : \sigma_1 \to \sigma_2$ takes a point at the first time to the unique point at the later time to which it is connected by a "geodesic". Thus, T_1 resembles GTR in that their models have determinate dynamic structure. In the case of GTR, for each spacetime M, and for each $p \in M$, there is a unique vector v in the tangent space over p that specifies the four-momentum of a particle at p.

We claim that the theory T_1 is genuinely deterministic, and only trivially semantically indeterministic. First to establish that T_1 is deterministic, we show that its models satisfy Property R. Suppose, for example, that M and N are models of T, and that $f, g : M \to N$ are isomorphisms such that $f_{\sigma_1} = g_{\sigma_1}$. By the definition of homomorphisms of models, the following diagram commutes

$$\begin{tikzcd}
M(\sigma_2) \arrow[r, shift left, "f_{\sigma_2}"] \arrow[r, shift right, "g_{\sigma_2}"'] & N(\sigma_2) \\
M(\sigma_1) \arrow[u, "M(\gamma)"] \arrow[r, shift left, "f_{\sigma_1}"] \arrow[r, shift right, "g_{\sigma_1}"'] & N(\sigma_1) \arrow[u, "N(\gamma)"']
\end{tikzcd}$$

and therefore $f = g$.

Despite being deterministic in an obvious sense, there is another sense — a trivial semantic sense — in which T_1 is indeterministic: the choice of domain sets $M(\sigma_1)$ and $M(\sigma_2)$ does not constrain the choice of the isomorphism $M(\gamma)$ that connects the two of them. Indeed, changing this isomorphism generates a "problem" for T_1 that is identical in structure to the hole problem of GTR. For example, let M be the model of T_1 with $M(\sigma_1) = \{a, b\}$, $M(\sigma_2) = \{a', b'\}$, and where $M(\gamma)$ is the function that takes a to a' and b to b'. Let N be the model of T_1 that is just like M except that $N(\gamma)$ is the function that takes a to b' and b to a'. Then the models M and N are isomorphic, as witnessed by the isomorphism $f : M \to N$ such that f_{σ_1} is the identity on $\{a, b\}$, and f_{σ_2} flips elements of $\{a', b'\}$.

Obviously, in the case of T_1, the indeterminism is trivial and semantic. But on the question of determinism, there is no relevant difference between T_1 and GTR, and hence the indeterminism of GTR is trivial and semantic.

For our second example, let $\Sigma_2 = \{\sigma_1, \sigma_2\}$, and let T_2 be the theory in Σ_2 that says that the two domains have the same cardinality. (This restriction on cardinality is purely to avoid distractions about dissimilarity with the first example.) Unlike T_1, this theory T_2 is genuinely indeterministic in the following sense: there are models M and N of T_2, and isomorphisms $f, g : M \to N$ such that $f_{\sigma_1} = g_{\sigma_1}$ but $f \neq g$. Indeed, any bijection $f_{\sigma_1} : M(\sigma_1) \to N(\sigma_1)$ can be combined with any bijection $f_{\sigma_2} : M(\sigma_2) \to N(\sigma_2)$ to give an isomorphism $f : M \to N$. This shows that models of T_2 do *not* have determinate dynamical structure that connects earlier states of affairs to later states of affairs.

Interestingly, both of these toy theories, T_1 and T_2, are prima facie substantivalist, for they quantify over spacetime points. If there is indeterminism in a spacetime theory, it does not come from a commitment to the existence of spacetime points, but from a lack of commitment to objective dynamical structure.

4 Metric essentialism

There are interesting parallels between our take on the hole problem and Maudlin's metric essentialism, according to which "spacetime points have their spatio-temporal properties essentially" (see Maudlin, 1990). Indeed, on a charitable reading, our insistence that the hole argument would need a non-trivial isometry is already implicit in Maudlin's claim that spacetime is represented by a metric space.

> 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 space-time is represented not by the bare manifold but by the manifold plus metric, by the metric space. (Maudlin, 1990, p 546)

Since Maudlin claims that spacetime is represented by a metric space, he would appear to agree with us that the correct notion of isomorphism of spacetimes is isometry. Nonetheless, Maudlin is mistaken to think that the reason that metric properties are invariant under isomorphism is because metric properties are essential properties of spacetime regions.

We will illustrate the issues with a simple example. Let $\Sigma = \{P\}$ be a signature for a first-order language, with P a unary predicate symbol, and let

T be a first-order theory in signature Σ. According to the standard textbook definition (Marker, 2006, p 9), an isomorphism between models M and N of T is a bijection $f : M \to N$ such that $f[P^M] = P^N$, where P^M is the extension of P in the model M. In other words, isomorphisms preserve the extension of predicates and relations.

What is the motivation for this way of defining isomorphism? The basic idea behind isomorphism is structural sameness, i.e. two isomorphic situations should display the same structural features, even if their material contents are very different. For example, if two situations are isomorphic, then there are exactly the same number of things in the extension of P in the first situation as there are in the extension of P in the second situation.

What is *not* required of isomorphic situations is that the same individuals appear in the two situations, or that those individuals have the same properties in the two situations. For example, the situation in which Alice is P and Bob is not-P is isomorphic to the situation in which Alice is not-P and Bob is P. In other words, the motivation for saying that isomorphism preserves the extension of P has nothing to do with individuals having their P properties essentially; it has to do with not caring which individuals instantiate P.

Things do get a little more confusing in the case of automorphisms, i.e. isomorphisms whose domain and range are the same model. In particular, an automorphism $f : M \to M$ does not change the extension of predicates such as P, i.e. $a \in P^M$ iff $f(a) \in P^M$. Nonetheless, we do not grant that this feature of automorphisms means that predicates like P represent modally rigid properties, i.e. properties that objects either necessarily have or necessarily lack. (That claim would prove way too much, because theories surely do make use of contingently possessed properties!) Indeed, an automorphism is a special case of an isomorphism, and so the motivation for the requirement that an isomorphism maps P^M to P^N must be more general than the motivation that an automorphism preserves P^M. Moreover, the motivation for the requirement that an isomorphism maps P^M to P^N cannot be that P represents a modally rigid property, because models M and N need not have any individuals in common.

The same idea applies to isomorphisms of relativistic spacetimes, which are, after all, models of GTR. The reason why isomorphisms should be isometries, i.e. preserve metric structure, is not because spacetime points bear their metric properties and relations essentially. Rather, the reason is that the theory does not care at all about the identity of the points that instantiate the metric structure; it cares only about which metric structure is instantiated.

To summarize, Maudlin suggests that isometry is the right notion of sameness between spacetimes, and on this issue, we agree. What's more, if isometry is the right notion of sameness of spacetimes, then the hole argument presupposes that there are non-trivial hole isometries. Since no such isometries exist, there is no mathematical basis for the hole argument.

5 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. Unfortunately, what it means to say that possible worlds are the same, or that they agree on an initial segment, is left vague and undefined.

Fortunately, philosophers of physics do not have to depend on their own ingenuity to explicate the notion of identity of possible worlds. Instead, they can use the same tools that physicists use when they reason about spaces of models of a theory, or spaces of solutions to differential equations — tools such as topology and category theory. That is, a theory, such as GTR, is not specified by some vague collection of possible worlds, but by a collection of models equipped with some well-defined mathematical structure.

In particular, a reasonable regimentation of GTR will treat spacetimes as isomorphic only if they are isometric. And in that case, GTR is provably deterministic, whether or not one thinks that spacetime is a substance.

References

Earman, John (1986). *A primer on determinism*. Springer.

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.

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

Halvorson, Hans (2019). *The logic in philosophy of science*. Cambridge University Press.

Lewis, David (1983). “New work for a theory of universals”. In: *Australasian Journal of Philosophy* 61.4, pp. 343–377.

Malament, David B (2012). *Topics in the foundations of general relativity and Newtonian gravitation theory*. University of Chicago Press.

Marker, David (2006). *Model theory: an introduction*. Springer.

Maudlin, Tim (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.

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

Weatherall, James Owen (2018). “Regarding the ‘hole argument’”. In: *The British Journal for the Philosophy of Science* 69.2, pp. 329–350.