The Hole Argument

John Norton

University of Pittsburgh

The "hole argument" shows that spacetime substantivalism leads to a radical form of indeterminism in a broad class of spacetime theories.[1] My purpose in this note is twofold. First I shall present a short and informal version of the argument in the form developed in Earman and Norton (1987). Second I will show how the argument can be extended to the case of "manifold plus further structure" substantivalism whenever that further structure admits certain common symmetries.

1. An Illustration of the Mathematical Devices Used

The hole argument depends on the possibility of displaying two models of some spacetime theory which agree everywhere but within a small neighborhood of the spacetime manifold. To illustrate how two such models are arrived at, I will describe the construction for the easy to visualize special case of a spatially homogeneous and isotropic expanding universe in general relativity. It will be clear how the construction can be extended to other spacetime theories.

A model of general relativity is a triple <M,g,T> which represents a physical situation deemed possible by the theory. M is a four dimensional differentiable manifold which is a set of point-events laid out in a continuum with neighborhoods that are four dimensional. For the above case of an expanding universe, the stress-energy tensor T represents the smoothed out matter of the galaxies. In Figure 1, this smoothed out matter is pictured by the world lines of the galaxies. The diverging of these world lines is a manifestation of the expansion of the universe. Each of these galaxies is in free fall. The metrical structure g of the spacetime determines which trajectories in the manifold are free fall trajectories as well as a large number of other properties related to gravitation and the metrical behavior of rods and clocks. This metrical structure is pictured in Figure 1 by the little light cones drawn at various places.

A diffeomorphism on the manifold M is just a map that assigns points in M to points in M in a smooth, invertible manner. We are interested in a special case which we call a "hole diffeomorphism". To define one, we choose any neighborhood of M we please and call the chosen neighborhood "The Hole".[2] A hole diffeomorphism h is just the identity map outside The Hole and comes smoothly to differ from the identity within The Hole. An example is shown in Figure 1. A diffeomorphism can "carry along" the structures g and T defined on the manifold. If the diffeomorphism h maps the point p to the point hp,

PSA 1988, Volume 2, pp. 56-64

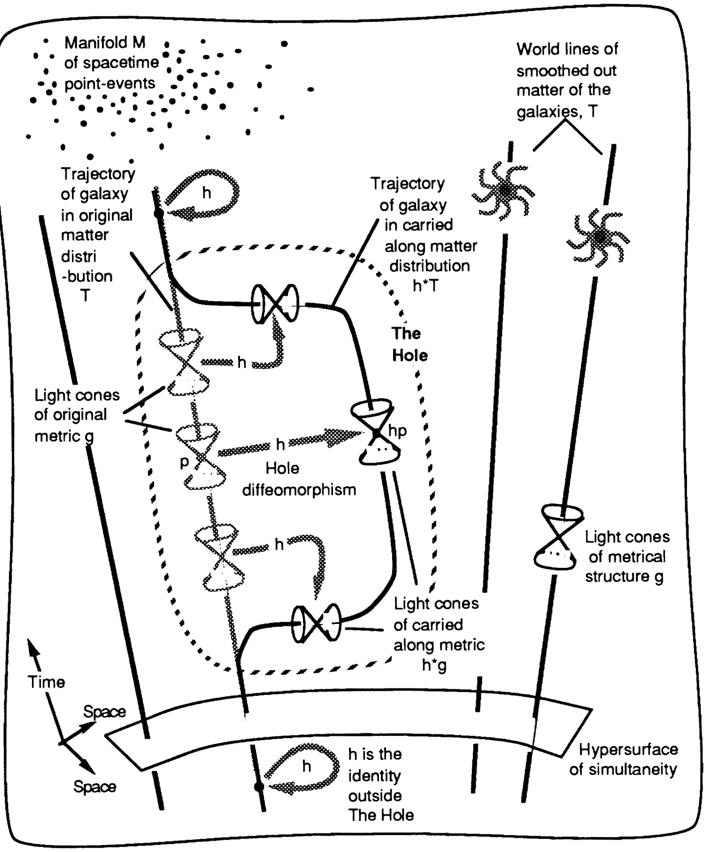

Figure 1 Hole Diffeomorphism Applied to an Expanding Universe in General Relativity

then a world line that passes through p is carried along to a world line that passes through hp. The carry along under hole diffeomorphism h of T is called h*T. It is shown in Figure 1. Notice that the carried along trajectories of the smoothed out matter of the galaxies are no longer free fall trajectories of the metric g. However we can also define the analogous carry along h*g of the metric g. The trajectories of the carried along galaxies will now be free fall trajectories of the carried along metric h*g.

We now have two triples: the original model <M,g,T> and a diffeomorphic copy <M,h*g,h*T>. The fact that the first triple is a model of the theory does not guarantee that the diffeomorphic copy is also a model, that is, also represents a physically possible situation according to the theory. A very important theorem in general relativity, however, assures us that diffeomorphic copies of models are themselves models. This theorem depends only on the fact that general relativity is what we call a "*local spacetime theory*": the fields g and T of the theory are specified solely by tensorial differential equations. Most commonly discussed spacetime theories can be formulated as local spacetime theories in which all the fields of the models are determined by tensorial differential equations. The hole argument applies to all spacetime theories in their local formulations. Newtonian spacetime theory and special relativity, for example, both admit local formulations[3] and the result is given generally as:

Gauge Theorem (Active General Covariance): If Mod is any model of a local spacetime theory and h any diffeomorphism defined on its manifold then the diffeomorphic copy h*Mod is also a model of the theory.

The standard assumption in general relativity texts is that two diffeomorphic models such as these represent the same physical situation. This assumption is an instance of the more general assumption which applies to all spacetime theories and is stated as:

Leibniz Equivalence: Diffeomorphic models in a spacetime theory represent the same physical situation.

A common justification for this equivalence is the fact that diffeomorphic models agree on all observables under the standard physical interpretations of the mathematical structures. For example, the original and carried along galaxies of Figure 1 may traverse The Hole by very different trajectories. But the time each requires to traverse the hole, as measured by co-moving clocks will be the same provided the first is determined by the original metric and the second by the carried along metric. Similarly any other observable that we might select pertaining to the phenomena in The Hole will fail to distinguish between the original and carried along model. A general justification of this claim and Leibniz Equivalence is based on Einstein's "point-coincidence" argument. (See Norton, 1987.) This argument asserts that all observables are fully reducible to coincidences at point-events such as the collision of two particles or the coincidence of a pointer with a given mark on a scale. Since all such coincidences are preserved under the carry along, all observables must also be left unchanged in the transition to the carried along model.

2. Consequences of the Denial of Leibniz Equivalence: Radical Local Indeterminism

Two diffeomorphic models are in general distinct mathematical entities—they will have different components in the same coordinate chart. Thus we are not forced by logical necessity to assume that both represent the *same* physical situation, although, of course, this assumption is routinely made in general relativity texts. Let us pursue the consequences of denying Leibniz equivalence.

If we deny Leibniz equivalence, we conclude that diffeomorphic models represent distinct physical situations. However diffeomorphic models agree on all observables. So we must conclude that they represent distinct physical situations which cannot be distinguished by any observationally verifiable differences. In the heyday of logical positivism this conclusion alone would have been sufficient to terminate any further consideration of the denial of Leibniz equivalence.

For those undeterred by observationally unverifiable differences, there is a further undesirable consequence revealed by the hole argument: radical local indeterminism in all local spacetime theories. To see how this indeterminism arises, take the example of the local spacetime theory of general relativity and consider the two models of an expanding universe pictured in Figure 1. The two models agree exactly everywhere outside The Hole since the diffeomorphism that carries one into the other is the identity outside The Hole. But they come smoothly to differ within The Hole. Thus the fullest specification of all the fields outside The Hole will not enable the theory to determine how the model will develop into The Hole, *no matter how small The Hole is in spatial and temporal extent*. This indeterminism is of a very extreme form rarely encountered in non-quantum theories. Under it, the only way that one can determine the model over the entire manifold is by specifying it everywhere. If the specification omits any neighborhood no matter how small, the above construction shows us that the theory will fail to determine the model within that neighborhood.

It is important to see that this form of indeterminism is undesirable because of the special way that it arises. Since diffeomorphic models agree on all observables, the denial of Leibniz equivalence amounts to the assumption that there are physically significant properties of the models that transcend observational verification. The construction of the hole argument reaffirms the dubious nature of these extra properties. It shows that these extra properties are not only opaque to observation but are also opaque to the theory itself in this sense: the theory is unable to determine how these properties will develop into an arbitrarily small neighborhood of the manifold, given a full specification of the properties everywhere else.

3. Who Must Deny Leibniz Equivalence?

Spacetime substantivalists must deny Leibniz equivalence. Spacetime substantivalism is an extreme form of realism about certain structures in spacetime theories. It holds that spacetime is a substance in so far as it is something that has an existence independent of its contents. I do not know how to make the notion of "independent existence" precise here. But this much is clear of the substantivalist position. If the contents of spacetime are rearranged in some way in spacetime—for example everything is spatially translated three feet in some direction—then the substantivalist must say that we arrive at a physically distinct situation. For an important physical property has changed: the spatiotemporal locations of the contents.

This necessary commitment of substantivalists and the ensuing conclusion of the physical distinctness of observationally indiscernible states of the world is precisely what Leibniz exploited in his challenge to the Newtonian Clarke in his third letter of their famous correspondence (Alexander, 1956). Leibniz considered, for example, the case of the bodies of the world replaced in space in such a way that East and West are exchanged but all other relations preserved. The Newtonian space substantivalist must insist that a new world has been formed even though it is indiscernible from the old one.

In local spacetime theories, the mathematical entity which most naturally represents spacetime are the manifolds of the models. This view is defended in Earman and Norton (1987). It leads the substantivalist to what we call "manifold substantivalism". The manifold substantivalist must insist that the rearrangement of spacetime structures against the background of the spacetime manifold leads to a structure that represents a different physical situation. Since the carry along under diffeomorphism effects just such a rearrangement, the manifold substantivalist must insist that a model of a spacetime theory and a (non-identical) diffeomorphic copy of it represent different physical situations. Thus the manifold substantivalist must deny Leibniz equivalence and accept the undesirable consequences outlined in Section 2. In particular the extra physical properties introduced by this substantivalist must be opaque to both observation and the laws of the physical theory in the sense given above.

4. An Escape? Manifold plus Further Structure Substantivalism

The full force of the hole argument is directed against manifold substantivalism. One might think, therefore, that the substantivalist can escape the undesirable consequences of the hole argument by choosing not just the manifold, but the manifold plus some further structure to represent spacetime. For example one might choose the manifold plus metric in both special and general relativity. Or in Newtonian spacetime theory one might choose the manifold plus absolute time one-form, Euclidean spatial metric and the affine connection adapted to them. For concreteness below, I shall assume that manifold plus further structure ("mpfs") substantivalists do make these choices for the "further structure" in Newtonian theory, special and general relativity.

A simple intuitive consideration reveals roughly when this escape via mpfs substantivalism will succeed or fail. The hole argument goes through against the manifold substantivalist because the spacetime theories we consider provide no means of individuating physically the points of the manifold other than through the further structures defined on them. Thus one produces no changes in the observational consequences by rearranging the individuating structures over the manifold by means of a carry along. Moreover the laws of the theory seem indifferent to whether one carries out such rearrangement. Thus we would expect similar problems for mpfs substantivalism if the "further structure" exhibits symmetries. For loosely speaking, the presence of these symmetries represent a failure of the further structure to individuate fully the points of the manifold. These intuitions are made more precise in two claims.

> *Claim 1: Observational indistinguishability.* If the "further structure" selected by the mpfs substantivalist admits any symmetry transformation at all, then the mpfs substantivalist is committed to the distinctness of observationally indistinguishable states of affairs in local spacetime theories.

Justification. Let the theory have models <M,S,C>, where M is a differentiable manifold, S represents the "further structure" and C represents the remaining "contents" of spacetime. Let S have the non-identical symmetry h, so that by definition h*S=S. Consider the two structures <M,S,C> and <M,S,h*C>. If one of them is a model of a local spacetime theory, the gauge theorem guarantees that both are models. The mpfs substantivalist must say that they represent physically distinct situations since C has been rearranged against the spacetime background of <M,S>. But we have that <M,S,h*C>=<M,h*S,h*C>. Thus the mpfs substantivalist must say that the two diffeomorphic models <M,S,C> and <M,h*S,h*C> represent physically distinct situations. However from earlier we know that such diffeomorphic models agree on all observables.

Theories that exhibit symmetries required for Claim 1 to hold include flat Newtonian spacetime theory, special relativity and general relativity applied to spatially homogeneous and isotropic cosmologies.[4]

> *Claim 2: Indeterminism.* If the "further structure" includes certain common symmetries, such as spatial homogeneity and isotropy, then we can recover a form of indeterminism analogous to that arrived at in the hole argument in local spacetime theories.

Justification. The construction needed to establish this result is more complicated than that required for the hole argument. I will give it for the case of special relativistic electrodynamics. It will be clear that analogous constructions are possible for the other theories cited. Special relativistic electrodynamics has models <M,g,F>, where M is an R^4 differentiable manifold, g a Lorentz signature metric tensor and F the Maxwell field tensor. The model set of the theory contains just all such triples in which g satisfies the vanishing of the Riemann curvature tensor as its tensorial field equation and F satisfies Maxwell's equations. For simplicity we assume that F is fully inhomogeneous and anisotropic. Assume that the mpfs substantivalist selects <M,g> as representing spacetime. Let t be a translation by unit spatial distance in some direction. t is a symmetry of g so that we have

t*g=g.

It is possible to decompose the translation t into the composition of two parts as shown in Figure 2. To do so, we select any two parallel hypersurfaces of simultaneity, which divides the manifold into three regions. Call the the one between the hypersurfaces the "present" and the remaining two the "past" and the "future".

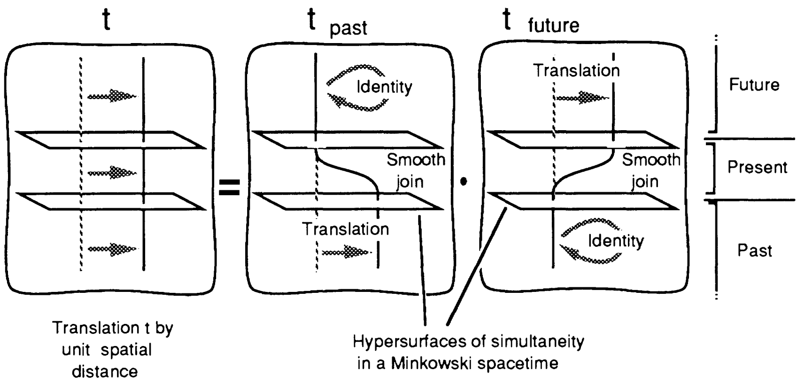

Figure 2 Decomposition of a Spatial Translation in a Minkowski Spacetime

Define t_{past} to be the identity map in the future, to coincide with t in the past and, in the present, to be a smooth interpolation between them. Similarly define t_{future} to be the identity in the past, to coincide with t in the future and, in the present, to be a smooth interpolation between them. The two interpolations in the present are to be chosen so that we have:

$t = t_{past} \cdot t_{future}$.

In accord with the earlier discussion, the mpfs substantivalism must hold that translating F to t*F across the spacetime background of <M,g> will produce a structure representing a physically different situation; the various parts of F are now located at different spatiotemporal events. That is, this substantivalist must conclude that the two models <M,g,F> and <M,g,t*F> represent different physical situations. Exploiting the fact that t*g=g we can say:

Model I, <M,g,F>, represents physical situation I.

Model II, <M,t*g,t*F>=<M,g,t*F>, represents physical situation II.

Now recall that we could proceed from model I to model II by a two step transformation: the application of t_{future} followed by the application of t_{past}. Figuratively:

<M,g,F> (t_{future})→ <M,t_{future}*g,t_{future}*F> (tpast)→ <M,t*g,t*F>=<M,g,t*F>

Let us call the intermediate model <M,t_{future}*g,t_{future}*F> "model III". Note that it is diffeomorphic to both model I and II, so it represents a physical situation observationally indistinguishable from those represented by models I and II. Now ask what physical situation is represented by model III. There are three possible answers:[5]

(a) physical situation I

(b) physical situation I

(c) a physical situation other than I or II.

Any answer leads to indeterminism. For example, consider answers (b) or (c). The selection of either entails that the two diffeomorphic models, model I and model III, represent distinct physical situations. However by the construction of the diffeomorphism t_{future} that relates the two models, they agree in the past and come smoothly to disagree in the future—and the disagreement matters physically since the models represent different physical situations. This is a classic instance of future directed indeterminism. The full specification of the model in the past fails to determine the development of the model into the future. Correspondingly, if the answer selected is (a), one considers model III and model II which leads by analogous arguments, to past directed indeterminism. (The time reversibility of the theory enables us to convert this past directed indeterminism into a future directed indeterminism if we wish.)

The construction given is one of the simplest. It could be extended to give the indeterminism more of a "hole-like" character. For example, by decomposing the translation t into more parts we could give a version of the construction in which the underdetermined neighborhood consists of only an arbitrarily thin slice of the manifold bounded by two parallel hypersurfaces of simultaneity.

Finally, I note that the above construction exploited only the fact that the spacetime <M,g> of special relativity admits spatial translations as symmetries. Thus analogous arguments can be mounted in any theory which admits such symmetries or analogous ones like like spatial rotations. Thus it follows that an indeterminism akin to that of the hole argument faces mpfs substantivalists in at least the following theories: flat Newtonian spacetime theory, special relativity and general relativity applied to spatially homogeneous and isotropic cosmologies.

5. Conclusion

The analysis of spacetime substantivalism given here is not based on the belief that determinism is or ought to be true. Earman (1986) has catalogued admirably the many ways that determinism can fail even in classical theories. The force of the attack on spacetime substantivalism comes from the way that the indeterminism arises. The spacetime substantivalist is forced to introduce properties which must have physical significance, even though they remain inaccessible to observational verification and are opaque to the laws of the spacetime theory in so far as the theory cannot determine their development.

Addendum: Replies to the Criticism of Maudlin and Butterfield

Tim Maudlin's essentialism offers an escape for substantivalists from the hole argument which depends on the analysis of which properties are essential to spacetime. In so far as the escape reduces to the endorsement of manifold plus further structure substantivalism,[6] I have outlined in Section 4 above the circumstances under which the escape fails and succeeds. Whenever the spacetime admits no symmetries, then the mpfs substantivalist escape succeeds. Whenever the spacetime admits symmetries in the way indicated, then the mpfs substantivalist escape fails. Overall we might say that the escape enjoys partial success. For, through it, spacetime substantivalism need not *always* lead to the disastrous consequences—they only arise sometimes. Of course there is a surer escape: the denial of spacetime substantivalism.

Unfortunately I do not think that Jeremy Butterfield's most ingenious escape via counterpart theory even enjoys this type of partial success. Butterfield's escape depends on endorsing what he calls "One", which asserts that at most one of two diffeomorphic models of a spacetime theory can represent a physically possible world. If "model" means "mathematical structure which a theory selects as representing a physically possible situation" then "One" ought to assert that at most one of two diffeomorphic *structures* can represent a physically possible world according to a given spacetime theory. The first problem is that "One" directly contradicts the active general covariance of our local spacetime theories as expressed in the gauge theorem above. The models of a local spacetime theory are *all* structures of the appropriate type that satisfy the theory's tensorial field equations. John Earman and I stressed this "all" in our (1987, p.517) by explicitly introducing a "completeness condition". This completeness condition allows derivation of the gauge theorem in local spacetime theories. This theorem contradicts "One" by guaranteeing that a diffeomorphic copy of a model is itself a model and thus represents a physically possible situation.

Let us set this worry aside. We might choose, for example, to deviate in some way from the local formulation of the spacetime theory. I still do not think that the escape works. Given two diffeomorphic structures, at most one is a model of the theory according to "One". Take the case in which one of them is a model. How are we to distinguish the real model from the imposter? There must be some property which distinguishes them and the property must be physically significant in so far as it tells us which structure represents a physically possible world. Since the real model and the imposter are diffeomorphic, this property cannot have observational consequences. If we mistakenly choose the imposter as a model, we would interpret it as representing a physically possible world indistinguishable observationally from that represented by the real model. Similarly the property eludes the tensorial field equations of local spacetime theories. Take exactly the set up of the hole argument with the real model specified on the manifold everywhere outside The Hole. These field equations will be unable to distinguish between the development into The Hole of the real model or of one of its infinitely many diffeomorphic copy-imposters. Thus the counterpart theorist is in precisely as bad—or as good—a situation as the manifold spacetime substantivalist. Both introduce properties which must have physical significance, even though they remain inaccessible to observational verification and are opaque to the laws of local spacetime theories.

Notice finally that the problem of the real model and the imposter must face anyone who would seek an escape from the hole argument by avoiding the local spacetime formulation of spacetime theories.

Notes

[1]The hole argument was advanced by Einstein in 1913-1914 as an argument against the acceptability of generally covariant field equations in general relativity. Its clearest statement is Einstein (1914, pp.1066-67) and the role it played in Einstein's thought is discussed in Norton (1987). The non-triviality of the argument was revealed to modern readers by Stachel (1980). For a novel approach to the reading of Einstein's version of the argument, see Norton (forthcoming).

[2]The name was introduced by Einstein for his version of the argument which applied to a special case in which The Hole was a matter free neighborhood—a hole!—in a matter distribution.

[3]For special relativity, the theory has models <M,g>, where M is a four dimensional differentiable manifold and g a Lorentz signature metric. The field equation is simply the vanishing of the Riemann curvature tensor.

[4]In both Claim 1 and 2, for the case of general relativistic cosmologies, the constructions must use a matter distribution whose *stress-energy tensor* is spatially homogenous and isotropic, but which is spatially inhomogeneous or anisotropic in some other property. An example of such a matter distribution is a uniformly expanding cloud of non-interacting dust particles, each with the same mass, but not all of them identical particles.

[5]One might be tempted to add the fourth answer "no physical situation at all". To do so violates the gauge theorem stated above for local spacetime theories. If one persists nonetheless, one must then face the problem of the real model and the imposter described in my reply to Butterfield below.

[6]A closer reading of Maudlin's text reveals to me that he cannot advocate the type of mpfs substantivalism described above. He holds that a hole diffeomorphism fails to transform a model of a theory into another model, since the transformed mathematical structure does not represent a physically possible situation. Thus his escape fails in the same manner as Butterfield's: first, he violates active general covariance and, second, he must face the problem of the real model and the imposter. (I thank Jeremy Butterfield for this point.

References

Alexander, H.G. (1956) (ed.), *The Leibniz -Clarke Correspondence.* Manchester Univ. Press.

Earman, J. (1986), *A Primer on Determinism.* Dordrecht: Reidel.

Earman, J. and Norton, J. (1987), "What Price Spacetime Substantivalism? The Hole Story", *British Journal for the Philosophy of Science* 38: 515-525.

Einstein, A. (1914), "Die formale Grundlage der allgemeinen Relativitaetstheorie", *Preuss. Akad. der Wiss., Sitz.* pp.1013-1085.

Norton, J. (1987), "Einstein, the Hole Argument and the Reality of Space", in *Measurement, Realism and Objectivity*, J.Forge (ed.). Dordrecht: Reidel, pp.153-188.

______. (forthcoming), "Coordinates and Covariance: Einstein's View of Spacetime and the Modern View", *Foundations of Physics.*

Stachel, J. (1980), "Einstein's Search for General Covariance". Paper read at the Ninth International Conference on General Relativity and Gravitation, Jena. Published in D.Howard and J.Stachel (eds.), *Einstein and the History of General Relativity: Einstein Studies*, Volume 1. Boston: Birkhauser, 1989 forthcoming.