Einstein's Hole Argument and its Legacy

Hanns Hagen Goetzke
Hughes Hall
University of Cambridge

2.7.2011

Contents

1 Introduction

Albert Einstein, after he completed the theory of special relativity in his *annus mirabilis* 1905, worked on a generally covariant theory of gravitation. In August 1912 Einstein returned to Zürich and began to collaborate with his old friend Marcel Grossmann, who was a professor of descriptive geometry. Grossmann helped him to understand tensor calculus which became the required mathematical background for the later theory of general relativity.

During their fruitful collaboration Einstein and Grossmann considered field equations based on the Ricci tensor since this tensor was the natural choice fulfilling the mathematical requirements that came along with general covariance. In 1913 they came extremely close to the final field equations of the year 1915; but they discarded their equations on the grounds that they could not yield the correct Newtonian limit.

Einstein then showed his enormous independence of mind: If some approach or argument did not work, he tried something else, if necessary even the contrary of his first approach or argument. Thus, not being able to find generally covariant field equations, he then changed his mind and came up with two fundamental arguments *against* general covariance - one of the essential principles he had so far assumed a good theory describing our world must have. The important and more general argument of those two is Einstein's hole argument, which will be discussed in this essay.

In short, the argument is as follows: Some field equations are assumed, which are satisfied by the stress-energy tensor and a metric tensor on the whole spacetime. In some region (the 'hole') the stress-energy tensor vanishes. Using general covariance, one can attribute two different metric fields to the hole region that are nevertheless identical everywhere outside the hole region. The hole argument concludes that general covariance prevents the stress-energy tensor from uniquely determining the metric tensor and thus the gravitational field.

Einstein and Grossmann published their new theory in 1913 (the *Entwurf* theory), which proposes non-generally-covariant field equations. In 1915 Einstein became more and more dissatisfied with the *Entwurf* theory, since it did not yield the right anomaly in the motion of Mercury. He also found a mistake in the last step of the derivation of the *Entwurf* field equations, which Einstein rejected in late 1915. After starting again from almost generally covariant field equations in November 1915, he could finally derive generally covariant field equations that reduced to the Newtonian limit. However, the history of how Einstein developed the final theory of general relativity shall not be the focus of this essay. A detailed history can be found in Norton's essay "How Einstein found his field equations: 1912-1915".

After Einstein finished his final version of the field equations he put the hole argument into a new perspective: the metric field can still not be uniquely determined by the general covariant field equations, if points of the spacetime manifold are taken as having an independent individuality. But therefore, this assumption has to be abandoned. Instead, spacetime points can only be distinguished by reference to the metric field (or maybe some other tensor fields, if the stress-energy tensor is not 0), and therefore they have no independent individuality. If one somehow removes the metric field, then nothing is left (and in particular no spacetime manifold with individual points would remain).

For a long time, Einstein's hole argument was seen as a rather simple mistake. Einstein and Grossmann are commonly accused of having forgotten that restrictions to coordinate systems were necessary in the recovery of the Newtonian limit. This is due to the fact that generally covariant field equations hold by definition in any coordinate system, whereas the theory of Newtonian gravitation only holds in certain coordinate systems. The limiting restrictions are additional relations called 'coordinate conditions' that must be satisfied by the solutions to the limit field equations. Einstein and Grossmann, it is supposed, had ignored their freedom to apply coordinate conditions. Furthermore, Einstein is supposed to have not recognized that in different coordinate systems a given physical instance is expressed in terms of different mathematical functions.

In 1980, Stachel was the first author who rejected the common view and argued that Einstein's hole argument is based on nontrivial concerns. Norton strengthens this view and gives evidence for his arguments by citing Einstein's notebooks from his time in Zürich (Norton 1984). He argues that Einstein knew about his freedom to apply coordinate conditions that would yield the Newtonian form. Einstein's concerns had a deeper reason, which is based on the relationship between the spacetime manifold and the gravitational field. Following Norton's reasoning, one can better understand Einstein's final reconciliation to the hole argument, after having developed his final version of the field equations, that is based on denying individuality of points.

Due to this new interpretation of the hole argument, a discussion about substantivalism in the context of general covariance arose. The centuries-old debate between substantivalists and relationists, having its roots in the debate between Newton and Leibniz, was renewed. Earman and Norton argued that general covariance will lead to radical indeterminism using the hole argument. They conclude that to avoid the failure of determinism, which should only fail for a reason of physics, one must therefore deny substantivalism (1987: 524).

However, in 1987 Butterfield analysed the definition of determinism that led to radical indeterminism. Based on the initial value problem of general relativity, he constructed a new definition, with the feature that substantivalism in combination with general covariance does *not* violate determinism. This alternative definition is equally compatible with the intuitive idea of determinism. In fact, in Earman and Norton's arguments against substantivalism no precise definition of determinism is stated. By using Butterfield's alternative definition of determinism, one can avoid the threat and have a deterministic generally covariant theory without denying substantivalism.

First, some mathematical definitions will be needed to discuss the hole argument, which will be explained in section 2. Secondly, substantivalism will be explained and some aspects of the debate between substantivalists and relationists mentioned. In section 4 Earman and Norton's threat to substantivalism will be explained. To see why determinism will be violated by general covariance, a first definition of determinism will be discussed. Afterwards, in section 5, an alternative definition of determinism will be developed and its consequences will be discussed. There is a technical and a philosophical aspect to the alternative definition. This essay will be focussed on the technical aspect of determinism. It will not describe the philosophical aspect in much detail. The interested reader is referred to Butterfield's essay "The Hole Truth" (1989).

2 The hole argument

To discuss Einstein's hole argument and its consequences, some mathematical definitions are needed.

To any spacetime theory there is a set of models such that each distinct model will validate the theory. A model $<\mathrm{M},\mathrm{O_i}>$ contains one manifold M and a collection of geometric objects $\mathrm{O_i}$. These geometric objects are fields on the manifold. In general these fields are not just scalar, vector and tensor fields but also connections, which in combination with a metric encode the spacetime structure. In classical spacetime theories there are a spatial and a temporal metric h and t as well as the 4-dimensional connection D (not unique since the metric is degenerate) which captures the idea of constancy of a vector field along a curve. In relativistic spacetimes there is a single metric g and a unique compatible 4-dimensional connection D (since g is non-degenerate). Matter fields are also geometric objects, that describe how matter is "spread" on the spacetime. The laws of the spacetime theory, which are partial differential equations, will be obeyed by the geometric objects contained in a model of the theory. Nevertheless the models differ from each other on initial values and boundary conditions. A theory is a set of models: $\mathcal{T} = \{\text{models} < \mathrm{M}, \mathrm{O_i} >\}$. Different models each represent physically possible worlds of the spacetime theory. Thus a theory is a set of possible worlds: it is a set of different ways the world could be like.

A diffeomorphism d on M is a smooth bijective (one-to-one and onto) map from M to M. This map d on points induces a map d^* on the geometric objects called the 'drag-along of d'. Any properties and relations of argument-points $\{p, q, r, ...\}$ according to geometric objects $\mathrm{O_i}$ will be the same properties and relations for the image-points $\{d(p), d(q), d(r), ...\}$ according to the dragged-along geometric objects $d^*(\mathrm{O_i})$. For example the distance according to a metric g between two points p and q is the same as the distance between $d(p)$ and $d(q)$ according to the dragged-along metric $d^*(g)$.

With a diffeomorphism d from M to M' one can compare different models $<$M,$\mathrm{O_i}$$>$ and $<$M',O'$_\mathrm{i}$$>$. In general, the dragged-along objects $d^*(\mathrm{O_i})$ need not coincide with the geometric objects O'$_\mathrm{i}$ on M'. But if they coincide in any region S, then using d one can speak of *agreement* of the two models on the physical state on the region S.

There are two different ways to understand general covariance: a passive and and active sense; though broadly speaking, they are equivalent. A physical theory is generally covariant in the passive sense if the form of the physical laws is invariant under an arbitrary smooth coordinate transformation. With the above definitions we can define an active version of general covariance in terms of models of a spacetime theory:

General Covariance (**GC**):
If $<$M,$\mathrm{O_i}$$>$ is a model of the spacetime theory, and there is a diffeomorphism d of M onto M, then $<$M,$d^*(\mathrm{O_i})$$>$ is also a model of the theory.

Not being able to find generally covariant field equations during his collaboration with Grossmann, Einstein thought about the reason why he was unsuccessful and wrongly concluded that there could be no general covariant field equations at all. His reasoning was based on the hole argument.

Einstein presented the hole argument four times between the years 1913 and 1914. The first three presentations are very brief. Norton argues that the hole argument being so condensed is the reason for the widespread misunderstanding of it which led to commentators accusing Einstein and Grossmann of having made a simple mistake (Norton 1984: 289). The argument for this focuses on the same topic as one of the two rival explanations, about why they abandoned the Ricci tensor: namely, they were not aware of their freedom to choose coordinate conditions. Following this assumption, one can explain away the apparent false conclusion in the hole argument. Thus Einstein is meant to have failed to recognize that choosing coordinate conditions does not change the physical content of the laws of the theory, but only alters their mathematical form in constraining them to special coordinate systems.

However, the fourth presentation of the hole argument is more detailed. The additional information given in this version is crucial for the understanding of the argument. Norton argues that his interpretation of the fourth presentation, which does *not* accuse Einstein and Grossmann of having made a simple mistake, can also be used for the three earlier versions (ibid: 287-290).

The hole argument states that if there are generally covariant field equations, then a given stress-energy tensor cannot uniquely determine the metric or the gravitational field only using the field equations. The argument is as follows: Einstein assumes that a stress-energy tensor and a metric tensor, which satisfy some field equations on the whole spacetime-manifold M, are given in such a way that the stress-energy tensor is 0 in a specific region H - called the 'hole'[1]. Einstein then argues that because of general covariance, two different metric fields can be assigned to the hole H, that satisfy the field equations and boundary conditions given by the tensors outside the hole. To show this, Einstein applies a change of coordinate systems: starting with the first coordinate system, which is defined on the whole manifold M (including the hole H) one can change the coordinates such that one gets a second coordinate system which coincides with the first system outside the hole H but is arbitrarily different inside the hole. This second system is used to build a second metric field. Einstein concludes that because of general covariance the stress-energy tensor can not uniquely determine the the metric tensor, i.e. the material content of spacetime can not uniquely determine the gravitational field.

[1]Einstein chose the expression 'hole', which is misleading in this context. The 'hole' does not describe a real hole in the manifold but a special region with certain properties. For a more intuitive understanding, the 'hole' argument should rather be called the 'patch' argument.

This radical indeterminism convinced Einstein that there are no generally covariant field equations.

It is a radical form of indeterminism for two reasons: the hole can have an arbitrary metric field since the choice of the coordinate system in the hole is arbitrary as long as the boundary conditions are satisfied. Also, the hole can be arbitrarily small. So even the most exhaustive description of the fields outside the hole cannot uniquely determine the fields inside the hole. This does not depend on the size of the hole region. So even a very weak form of determinism, which assumes the fields of every region except for an arbitrarily small hole are given, fails.

Norton argues that the change of coordinate systems used by Einstein does indeed construct a different metric field (Norton 1984: 288-289). The change of coordinate systems induces a related active diffeomorphism. Having this diffeomorphism one can use it to induce a drag-along map which gives a second metric field. To be able to discuss (in section 4 and 5) Butterfield's solution of the threat to substantivalism by Earman and Norton, this important construction of the second metric field will be developed in a manner similar to Butterfield's notation (Butterfield 1987: 22):

To start, one needs two different coordinate charts $\{x\}$ and $\{x'\}$ that have the same open domain D which is a subset of the spacetime-manifold M. The hole H is an open region and the closure of the hole is contained in the domain D. We take the two coordinate charts to be identical outside the hole, i.e. $x(p) = x'(p)$. Define a diffeomorphism d from D onto D which sends a point p to a point q. The point q is to have the same numbers in the coordinate chart $\{x'\}$ as the point p has in the coordinate chart $\{x\}$. This is equivalent to defining d by $x'(d(p)) = x(p)$. Therefore d is the identity outside the hole but not the identity inside, i.e. $d|_{\text{D-H}} = id|_{\text{D-H}}$ and $d|_{\text{H}} \neq id|_{\text{H}}$. The drag-along d^* on geometric objects is induced by d. Since the metric tensor g is a geometric object it is dragged-along to d^*g. The two metric tensor fields g and d^*g are different, because points on D have metric relations according to d^*g which their pre-images have according to g. This is illustrated in figure 1.

The point p in figure 1 lies outside the hole, so the diffeomorphism d sends it to the identical point p. Inside the hole H, d sends q to r, r to s and similarly for the other points. Because of the general covariance (GC) both models are a solution to the field equations of the theory. But they are different, as r is at distance 2 from p in the first model, whereas in the second model the distinct point s is at distance 2 from p and therefore has the property of its pre-image.

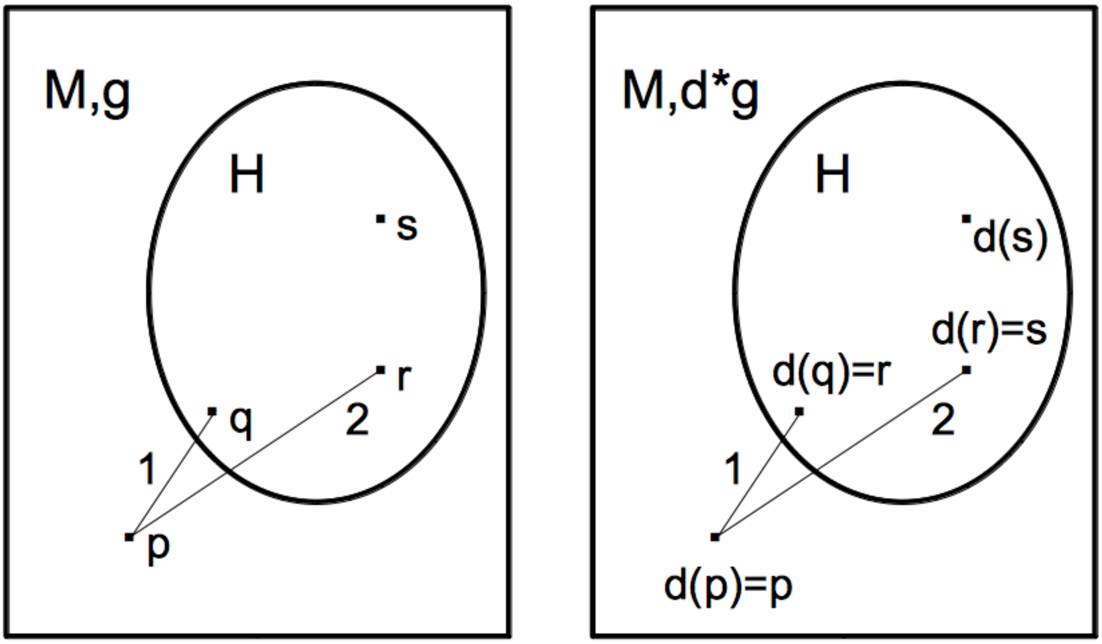

Figure 1: Points on a manifold M (left) are smoothly moved by a diffeomorphism d (right) within the hole H but are identical outside the hole.

Now recall that Einstein created the hole argument to show that general covariance leads to indeterminism, and therefore the theory of general relativity should not be generally covariant. Relying on this argument, Einstein was thus content that the *Entwurf* theory by Einstein and Grossmann was not generally covariant. However, one step in the derivation was not correct and the theory could not correctly predict Mercury's anomaly. In 1915, when he found the mistake in the derivation, Einstein got more and more dissatisfied with the *Entwurf* theory and started again to formulate a theory of general relativity. This time he was successful in predicting the anomaly of Mercury's motion. Furthermore, the field equations of the theory were generally covariant. Having found this result, he had to think again about his hole argument, since it contradicted his new correct result.

In a letter to Besso in January 1916 Einstein described his new understanding of the hole argument: The two models represent the same physical reality, since reality is only the sum of spacetime point-coincidences. These spacetime point-coincidences are conserved under a hole diffeomorphism d. Einstein writes that "it is most natural to require of laws that they determine no more than the totality of timespace coincidences. (...) This is already achieved with generally covariant equations" (Norton 1984: 291).

Einstein later used the hole argument to emphasize a different interpretation of the nature of spacetime: that points on the spacetime manifold have no independent individuality and can only be distinguished by reference to a metric field. Thus a spacetime manifold cannot exist without a metric field. So if one somehow removes the gravitational metric field from the manifold, then nothing remains. Einstein later wrote "Spacetime does not claim an existence of its own, but only as a structural quality of the field" (Butterfield 1987: 23).

To sum up: in the beginning of his search for a general theory of relativity, the principle of general covariance was very important for Einstein. And in the final stages it was crucial in his struggles towards the final field equations. But in 1913, his hole argument made him abandon it temporarily.

Shortly after he published his final theory, in 1917 Kretschmann argued that the requirement of general covariance does not make an assertion about the content of the theory as every spacetime theory can be expressed in a generally covariant way (Anderson, 1967: 338). Einstein accepted this objection. However, he probably assumed there was some deeper meaning of the general covariance principle. For him, the principle of general covariance was also a symmetry requirement. Today, in general relativity the general covariance is viewed as a gauge symmetry freedom.

After this introduction of the hole argument, some aspects of its legacy will be discussed in the next sections. In the following section substantivalism will be introduced so as to understand the threat to substantivalism by Earman and Norton.

3 Substantivalism

Ever since Newton published his theory of gravitation, there has been a grand debate between absolute versus relational theories of space and time (nowadays: spacetime).

It is hard to precisely define substantivalism as such, especially as its meaning varies between authors. Philosophers of space and time distinguish between several different versions of both substantivalism and relationism. Nevertheless, one can approach a definition and strengthen it by comparing it to relationism and describing the consequences of a substantival nature of spacetime. In this section substantivalism will be described and the rival theory, relationism, will be introduced so as to compare between the two doctrines. The approach to defining substantivalism will start from a scientific realist's point of view.

Scientific realism is a doctrine in the philosophy of science. It is the doctrine that the best available theories are approximately true. For a scientific realist, what those best theories postulate as existing does indeed exist.

If one applies scientific realism to general relativity, which is our best available theory of gravitation and electromagnetism in spacetime, one has to analyse the ideas behind the theory. Nowadays (and at the time Earman and Norton were first writing), most books about general relativity introduced the idea of spacetime as a manifold of points[2]. No further constraints on the nature of spacetime are imposed. Substantivalism is the doctrine that spacetime points are genuine objects. If spacetime points are objects, then regions, which are sets of points or aggregates of the points they contain, are objects, too. Spacetime itself is then the mereological fusion of all the points. A scientific realist, who takes general relativity (or other spacetime theories) seriously will therefore be inclined to endorse substantivalism.

Before the theory of general relativity was established, a substantivalist would say that spacetime is represented by a manifold and some additional geometric structure, called its absolute structure. In Newtonian or special relativity theories, which are formulated in a non-local way, a global absolute structure is defined *ab initio*, i.e. not as a physical field emerging from field equations as in general relativity. For special relativity, one could give the tuple $<\mathrm{N}, \eta_{\mathrm{ab}}>$ as a representation of spacetime.

The success of general relativity seemed to vindicate the relationist's point of view, since there is no *ab initio* global structure any more. It thus seemed that Newton's substantivalism was wrong, and Leibniz relationism was vindicated, after having been overshadowed by the success of Newton's substantival theory. But this common belief is misleading. In fact, there are only three respects in which general relativity has vindicated relationism (Butterfield 1987: 11-12).

The first vindication is based on the use of modern differential geometry to describe spacetime. In any classical or relativistic theory a spacetime is required to have a 4-dimensional connection to distinguish between accelerated and unaccelerated motion independent of a coordinate system. Not knowing modern geometry, it was natural for Newton to posit an absolute rest as this was the only way to define a connection. How-

[2]Maybe equipped with extra structure: e.g. a 4-dimensional Lorentzian manifold equipped with the Levi-Civita connection.

ever, with the use of modern differential geometry, the concept of absolute acceleration can be achieved by a connection without positing absolute rest, for both classical and relativistic spacetimes. Therefore, relationism if understood as denying an absolute rest is vindicated.

The second vindication is due to special and general relativity. In classical spacetime theories relationism fails, if it is defined as the assertion that motion can be described by the relational metric structure only. This is due to the fact that the 4-dimensional connection D is not uniquely determined by being compatible with spatial and temporal metrics h and t, i.e. $\mathrm{Dh} = \mathrm{Dt} = 0$. For there are different connections that are compatible with one pair of h and t. In relativistic theories, however, the metric g is non-degenerate. The connection is determined by the definition $\mathrm{D}g = 0$ and thus only the metric structure is needed to treat motion. Therefore relationism defined as claiming that motion is described sufficiently using only the metric structure is vindicated.

The third vindication is due to the interaction in general relativity between matter and geometric structure: they are influenced by each other. So the structure of the spacetime is not independent of the content. So relationism is vindicated if understood as denying spacetime as a fixed background for physical events independent of the distribution of matter.

One should note that these vindications of relationism should not be understood as supporting Mach's principle, which fails in both classical and relativistic spacetimes. Mach's principle states that only relative motion between bodies exists. This is not implied by the second and third vindication of relationism.

Even though relationism is vindicated in these three respects, there are good reasons to take a substantivalist's point of view; and in section 5 we will defend it against the threat formulated by Earman and Norton.

Some aspects of substantivalism are widely accepted. Of course, the absolute rest of Newtonian spacetime is not accepted. But the denial of Mach's principle is definitely accepted by substantivalists. So is the denial of *Leibniz equivalence*. Leibniz equivalence in terms of models is the claim that diffeomorphic models (two models are diffeomorphic if the geometric objects of one are the drag-along by a diffeomorphism of the geometric objects of the other) represent the same physical situation. A good example for Leibniz equivalence is the question whether there would be a different world if everything in the world were translated ten inches West, conserving all the relations between every object in the world. I shall take it that a substantivalist has to answer this question with a Yes

(Earman and Norton 1987:521-522).

In the following section, the threat to substantivalism pressed by Earman and Norton will be discussed. To follow their argument, their definition of substantivalism is adopted. For now we shall forget Einstein's later interpretation of the hole argument as denying independent individuality of the spacetime points, saying that they can only be distinguished by reference to a metric field.

Earman and Norton take substantivalism as the view that the bare manifold M of the models represent spacetime. They argue that all the geometric structure on the manifold is defined by fields, which are determined by partial differential equations. Therefore the metric and the connection are not inherent parts of the spacetime but physical fields in it. Even though the metric field encodes information about spatial distances and elapsed times, one excludes it from the spacetime. As the metric tensor comprises the gravitational field and therefore carries energy and momentum, it is classified as a content of the spacetime and not part of the spacetime itself (ibid: 518-519).

4 A threat to substantivalism

In 1987 Earman and Norton argued that general covariant field equations and substantivalism lead to radical indeterminism. They use Einstein's hole argument to develop their dilemma that one either has to give up substantivalism or accept a radical failure of determinism. But Earman and Norton's threat lacks an exact definition of determinism. Butterfield's replies to the threat provide precise definitions for different forms of determinism: of which one will be used in this section. First, an intuitive meaning of determinism will be introduced, before a precise definition in terms of models is stated.

Determinism is a property of a physical theory. An intuitive understanding of determinism is the Laplacian idea: a present state is the result of its past and the cause of its future. Knowing everything about the present state on a time slice S (or in a weaker version knowing everything about the present state on a time slice S and everything about its past) one could predict the state everywhere in the future. In other words, determinism says that a single physically possible world is uniquely identified by the physical state on a specific region of spacetime, i.e. given the state on that specific region, there is exactly one physical possibility for it (Butterfield 1987: 25).

In terms of models this means that if two models agree on a state at one time (or at one time and in the past), they will also agree in the future. But there is no immediate meaning of 'sameness of value' in two different manifolds. For example what does it mean for the vector at a point p in one model to be the 'same' as the vector at another point in another model? This comparison can be made using a diffeomorphism, and the drag-along map it induces on geometric objects, as described in section 2.

In general relativity not all the models have diffeomorphic manifolds, and therefore they cannot all be compared globally by one diffeomorphism. It is thus natural to make the definition of determinism conditional on the existence of a global diffeomorphism d between the models of the manifolds. The definition is also conditional on the existence of a region S of an appropriate kind $\mathcal{S}$. If those two conditions are not satisfied, then determinism will be true, but vacuously. Thus a first definition of determinism in terms of models, call it Dm1, is as follows (Butterfield 1987: 14):

Determinism 1 (**Dm1**):
Let $<\mathrm{M}, \mathrm{O_i}>$ be the models of a theory and let $\mathcal{S}$ be a kind of region occurring in manifolds of the kind that occurs in the models. A theory is then $\mathcal{S}$-deterministic, if and only if: given any two models $<\mathrm{M}, \mathrm{O_i}>$ and $<\mathrm{M'}, \mathrm{O'_i}>$ and any diffeomorphism d from M onto M', and any region $\mathrm{S} \subset \mathrm{M}$ of the kind $\mathcal{S}$:
If $d(\mathrm{S})$ is of kind $\mathcal{S}$ and $d^*(\mathrm{O_i}) = \mathrm{O'_i}$ on $d(\mathrm{S})$, which is of kind $\mathcal{S}$ too: then $d^*(\mathrm{O_i}) = \mathrm{O'_i}$ on all of M'.

Note that S is some subset of M and not necessarily a submanifold (e.g. a slice is not a submanifold). S is some preferred kind of region and the image under d of this region is of the same preferred kind. Also note that in contrast to the intuitive understanding of determinism the state on a region determines the state everywhere in the whole history. Not only the future development is determined, but also the past.

This version of determinism will be violated if a substantivalist assumes general covariant field equations: if one takes a hole diffeomorphism d (a diffeomorphism of the kind used in the hole argument in section 2), then the generally covariant field equations are not sufficient to determine all of the spatio-temporal properties a substantivalist is committed to. To see this, take two models $<\mathrm{M}, \mathrm{O_i}>$ and $<\mathrm{M'}, \mathrm{O'_i}>$ and a hole diffeomorphism d on a domain D. Let the domain D contain a hole region H. Then d is $d|_{\text{D-H}} = id|_{\text{D-H}}$ and $d|_{\mathrm{H}} \neq id|_{\mathrm{H}}$. Take the identity map id on M with $id^*(\mathrm{O_i}) = \mathrm{O'_i}$ on S where S = D - H. It is then obvious that $id^*(\mathrm{O_i}) \neq \mathrm{O'_i}$ on H. Because there are no absolute objects that d needs to preserve under the drag-along, the diffeomorphism d

is an example where Dm1 is violated for generally covariant field equations. And it is violated in a radical way as discussed above in section 2 (page 8).

Note that the threat cannot be avoided by assuming determinism to require only solutions that are unique up to an isomorphism. This is explicit in the example above. Dm1 makes solutions only unique up to an isomorphism but is violated by the above pair of models.

The threat is based on the denial of Leibniz equivalence. This leads to what Earman and Norton call the verificationist dilemma. It is not possible to observe spatio-temporal positions by themselves, because observables are just a subset of relations between structures on a spacetime manifold. Substantivalists must therefore either deny the Leibniz equivalence or accept that they cannot distinguish distinct state of affairs by observations (Earman and Norton 1987: 522). Thus a substantivalist must make distinctions that physics does not see. Earman and Norton then argue that the hole argument leads to radical indeterminism as sketched on page 8 since the field equations cannot uniquely determine the unobservably distinct states in the hole. The threat results in either the acceptance of radical local indeterminism, or the acceptance of Leibniz equivalence which is equivalent to denying substantivalism.

Earman and Norton's threat is not based on the conviction that determinism is true or should be true. They admit that there are many examples of how determinism can fail in spacetime theories, in classical Newtonian physics (e.g. space invaders from spatial infinity) as well as in general relativity (e.g. non-existence of a Cauchy surface). Also, determinism fails in quantum theory. (However, there are controversies whether the measurement process is really indeterministic, but this will not be important in this essay). Thus there is no general need to vindicate determinism. Determinism might fail. But if it fails it should not fail because of a philosophical doctrine but for a reason of physics (ibid: 524). In Short: determinism should not be violated in a radical way as a result solely of some general philosophical argument.

In the next section a new definition of determinism will be proposed so as to reconcile substantivalism with general covariance. Then determinism will *not* be violated by generally covariant theories like general relativity, and thus one can escape from the threat.

5 An alternative definition of determinism

Three responses to the threat to substantivalism are possible: one can give up determinism, one can give up substantivalism, or one can deny the threat and reconcile substantivalism and determinism. As above, determinism should only fail for a reason of physics and not because of some general philosophical doctrine. Earman and Norton were willing to give up substantivalism and they went on to discuss a way to formulate spacetime theories without quantifying over points. Using a result of Geroch, only one initial reference to points is needed in order to define geometric objects on a manifold: afterwards one can refer to the algebra of smooth scalar fields. If one takes the algebra as the fundamental object, representations would only be unique up to an isomorphism of spacetime models $<M, O_i>$. One can then define determinism as requiring that there be a subalgebra that determines the algebra; and so one can avoid the threat. This may very well be a feasible approach. But it is not necessary to either give up determinism or substantivalism: it is possible to reconcile the two (Butterfield 1987: 24-25).

To reconcile substantivalism with determinism, an alternative definition of determinism (Dm2) will be needed. There are two aspects to a new definition of determinism: a technical aspect and a philosophical aspect. Butterfield's proposed new definition is not just made up without a proper theoretical background. It is derived from the initial value formulation of general relativity, which will be introduced in this section. One can then see that for many well-known examples of spacetime theories, Dm2 will yield the same verdicts about whether the theory is deterministic as Dm 1 yields. However, there will be an example where Dm2 rules a theory as deterministic that Dm1 does not rule as deterministic. This is the technical aspect of the new definition.

The philosophical aspect concerns the question, whether Dm2 satisfies the intuitive idea of determinism as adequately as Dm1 does. Butterfield argues that Dm2 is faithful to the intuitive meaning of determinism; thus the technical success will not be hollow. Another concern of the philosophical aspect is whether Einstein's idea of distinguishing spacetime points by reference to a field is compatible with Dm2, whether it is necessary for the reconciliation and, going even further, whether it is right. This essay will not be focussed on the philosophical aspect of Dm2, but the main ideas will be introduced briefly. For a more detailed discussion the interested reader is referred to Butterfield's article "The Hole Truth" (1989).

In section 4 it was shown how Dm1 is violated by generally covariant field equations if one endorses substantivalism. But this result should not be automatically understood as a proof that any generally covariant theory is indeterministic, since general relativity textbooks prove uniqueness of solutions, if an appropriate initial value formulation is used. The uniqueness of solutions strongly indicates that general relativity satisfies some intuitive understanding of determinism. If one has control over initial conditions of a system in classical physics, then its behaviour can be completely determined if the system evolves in an isolated way, i.e. without interference from the outside. In practice it is very hard to control initial conditions for gravitational problems, but in principle it should be possible to control the initial conditions of the matter-distribution and the gravitational field in a region much smaller than the cosmological scale. Assuming that general relativity is not very different from other theories of classical physics, it is thus physically reasonable that the initial data can be specified. Therefore with this initial data, Einstein's equation should determine the evolution of the region (Wald 1984: 243).

A theory has an initial value formulation, if initial data can be specified and the dynamical evolution of the system is uniquely determined by this data. An initial value formulation is well posed, if it satisfies two conditions. The first condition concerns the intuitive understanding of determinism: small changes in the initial data should only result in small changes in the solution over a compact region. Initial conditions can only be measured to finite accuracy, so without this condition the theory would lose its predictive power. The second condition is that changes in the initial data in a region S will only affect the solution inside the causal future of S. Otherwise signals could be transmitted faster than the speed of light, which violates the principles of relativity theory. Wald argues that general relativity has such a well posed initial value formulation (ibid: 244).

To present the initial value formulation, some definitions and restrictions to the spacetime manifold are needed. If S is a region of a manifold M, the *future domain of dependence* is $\mathrm{D}^+(\mathrm{S})$ and the *past domain of dependence* is $\mathrm{D}^-(\mathrm{S})$. A *Cauchy surface* Σ is a region of M such that $\mathrm{M} = \mathrm{D}^-(\Sigma) \cup \mathrm{D}^+(\Sigma)$. This means that Σ is a region of spacetime that is intersected by every non-spacelike causal curve exactly once. In general relativity there is no single manifold, and the spacetime manifold's global topology and causal and temporal structure can be of a peculiar kind. To achieve an initial value formulation, it is assumed that the spacetime is *globally hyperbolic* (ibid: 255). A spacetime $(\mathrm{M}, g_{\mathrm{ab}})$ is globally hyperbolic, if it possesses a Cauchy surface (ibid: 201).

The initial data can be expressed in an intrinsic form without reference to a spacetime (M, g_{ab}) that the initial data is embedded in. The initial data consists of a triple $(\Sigma, \mathrm{h}_{\mathrm{ab}}, \mathrm{K}_{\mathrm{ab}})$, where Σ is a three-dimensional manifold, h_{ab} is a Riemannian metric on Σ and K_{ab} is a symmetric tensor field on Σ, that describes how Σ is embedded in M. The idea is then that given such initial data there exists a globally hyperbolic spacetime (M, g_{ab}), that satisfies Einstein's equations and contains a Cauchy surface diffeomorphic to Σ, and for which g_{ab} on this Cauchy surface induces the metric h_{ab} on Σ and the induced extrinsic curvature on Σ is K_{ab} (ibid: 256). Thus we take $\theta : \Sigma \to \mathrm{M}$ as the diffeomorphism such that $\theta(\Sigma)$ is a Cauchy surface in M. Then we say that the triple $(\mathrm{M}, \theta, g_{\mathrm{ab}})$ is called a *development* of $(\Sigma, \mathrm{h}_{\mathrm{ab}}, \mathrm{K}_{\mathrm{ab}})$.

Note that $(\mathrm{M}, \theta, g_{\mathrm{ab}})$ is not unique: a development can always be the extension of another and a development can be produced by dragging along another one with a hole diffeomorphism that leaves Σ fixed. However, the freedom to apply a hole diffeomorphism is included in the notion of an *extension.* An extension of a development is another development, which contains a subset that is isometric to the original development. But the converse inclusion is not correct. A development can extend a second one without using a hole diffeomorphism. The first can either be non-isomorphic to the second, or it can have a different base-set of points. But there exists a *maximal development*, which is an extension of every other development and is unique up to a diffeomorphism. Two maximal developments extend each other. (Butterfield 1987: 27).

These definitions and associated results will lead us to a notion of determinism appropriate for general relativity. To conclude the exposition the results, let us assume models, (M, g_{ab}) and (M', g_{ab}'). Then we have: if there is an initial data triple $(\Sigma, \mathrm{h}_{\mathrm{ab}}, \mathrm{K}_{\mathrm{ab}})$ and diffeomorphisms θ and θ' such that $(\mathrm{M}, \theta, g_{\mathrm{ab}})$ and $(\mathrm{M}', \theta', g_{\mathrm{ab}}')$ are maximal developments of $(\Sigma, \mathrm{h}_{\mathrm{ab}}, \mathrm{K}_{\mathrm{ab}})$, then (M, g_{ab}) and (M', g_{ab}') are isomorphic by an isomorphism that preserves Σ, i.e. $\theta(\Sigma)$ is mapped to $\theta'(\Sigma)$ (ibid: 28).

This looks similar to the definition of Dm1. But there are two main differences in the antecedents. First, because there is no reference of a global diffeomorphism in the antecedent, which gives a matching on Σ, it has to be claimed in the consequent that a global isomorphism exists. Secondly, the models are maximal and thus inextendible. In general relativity this is a common assumption to make.

With this, an alternative general definition of determinism can be presented. It will be expressed in terms of general models $<\mathrm{M},\mathrm{O}_{\mathrm{i}}>$, and Cauchy surfaces cannot be assumed because they do not exist in classical spacetimes. Instead, regions S of an appropriate kind $\mathcal{S}$ will be used, just like in section 4's discussion of Dm1 (ibid: 28-29):

Determinism 2 (**Dm2**):

Let <M, O_i> be the models of a theory and let $\mathcal{S}$ be a kind of region occurring in manifolds of the kind that occurs in the models. A theory is then $\mathcal{S}$-deterministic, if and only if: given any two models <M, O_i> and <M', O'_i>, which contain regions S and S' of kind $\mathcal{S}$ respectively, and any diffeomorphism α from S onto S':

If $\alpha^*(O_i) = O_i$' on $\alpha(S) = S$', then: there is an isomorphism β from M onto M' which sends S to S', i.e. $\beta(S) = S$', and for which $\beta^*(O_i) = O_i$' on all of M'.

This definition is different from Dm1: contrary to Dm1, the diffeomorphism α does not need to be a global diffeomorphism, it is sufficient that it be defined on S only. Also, the global isomorphism β does not need to extend α, i.e. it does not need to coincide with α on S. Therefore Dm2 does not reduce to Dm1, even if α is given globally. Thus Dm2 is a weaker definition of determinism than Dm1 is. However, since β does not need to extend α, Dm2 is not violated by a hole diffeomorphism. Thus, in principle, it allows generally covariant theories to be deterministic. Butterfield claims that this definition, which was derived by analysing general relativity's initial value formulation carefully, is the proper definition for generally covariant theories.

There is another aspect of Dm2, besides allowing generally covariant field equations to be deterministic from a substantivalist's point of view, which makes it an attractive definition of determinism. For most familiar spacetime theories, that assume a single manifold with a structure, Dm2 will decide whether the theory is deterministic or not in the same way as Dm1 does.

As an example, for theories using classical or Minkowski spacetime (globally $\mathbb{R}^4$), determinism will be analysed for the two different definitions, where either a slice or a sandwich is considered for determining the whole history. If α from the definition of Dm2 is a diffeomorphism between a slice or a sandwich on $\mathbb{R}^4$, then it can easily be extended to a global diffeomorphism on $\mathbb{R}^4$. It is then obvious that Dm1 and Dm2 will only differ by Dm2 being a weaker definition than Dm1, since β and α do not have to coincide. Thus for classical and Minkowski spacetimes if Dm1 rules a theory as deterministic, Dm2 rules it as deterministic, too.

The reverse is more interesting: Is there a spacetime theory that Dm1 rules as indeterministic whereas Dm2, being a weaker definition, rules it as deterministic? The answer is Yes. However, the example for this is a rather less familiar spacetime: Leibnizean spacetime is a classical spacetime, which has neither an absolute rest nor a connection.

Thus the models are <M, h, t>, where M is globally $\mathbb{R}^4$, and h and t are the spatial and temporal metric as before. This spacetime has little structure and therefore many symmetries. The symmetries which are important for the difference between Dm1 and Dm2 are, such that they are the identity map upto the time $t = 0$ and differ smoothly from the identity afterwards. Independent on the specific laws of the theory, such a symmetry will map a model containing matter to a different model. Therefore Dm1 fails, because if two models with the same manifold match by the identity map for $t \leq 0$, they will not match by the identity map for $t > 0$ and thus Dm1 is violated. On the other hand, Dm2 does not need to fail since the global match does not need to be the same map, that initially matched the two models on the region S (here the region S is the entire past before $t = 0$ and the initially matching map is the identity). For Dm2 the global match can be given by the symmetry itself. Nevertheless, this example is a rather special case and for most well-known spacetime theories Dm1 and Dm2 give the same verdicts about whether the theory is deterministic.

All this technical success of the new definition would be in vain, if Dm2 did not represent the intuitive idea of determinism as faithfully as Dm1 does. But Butterfield argues that both definitions satisfy the intuitive idea equally faithful (ibid: 25).

Furthermore, he argues that Einstein's idea of reconciling general covariance with determinism (using the idea that points are distinguished with reference to a field: cf. section 1 p. 4 and section 2 p. 10) is compatible with the alternative definition of determinism. It is, however, not a necessary part of the reconciliation.

To discuss these assertions, take the two models from figure 1. Substantivalism must deny that those two models represent the same physically possible world. This is because a substantivalist believes that the points and their relations and properties are fixed by a possible world. Thus different models cannot represent the same world. However, the denial can be of two different forms: first, one can take *each* model as representing a possible world. Then the two models violate the intuitive idea of determinism that a single physically possible world is uniquely determined by a state on a specific region of spacetime. But secondly, one can take *at most one* model as representing a possible world. This approach does not threaten the intuitive idea of determinism (ibid: 26).

There are again two ways to justify the second form of the denial: one can follow Einstein's approach, which is to distinguish the points by reference to a field. Some models will not represent a possible world, because they fail to assign the right properties and relations of some of the points. But Butterfield argues that the second way is philosophically attractive: it is a way which denies that one object can be in two possible worlds. In this case, it means that no point can be in two different worlds. Thus only one of the two models in figure 1 can represent a physically possible world, because one possible world could not share points with another one. Therefore Einstein's approach is not necessary to save determinism if one believes in substantivalism (ibid: 26).

6 Conclusion

On his way to developing a theory of general relativity, Einstein stumbled over the hole argument. It took him two years to find the mistake in Einstein and Grossmann's non-generally covariant field equations of general relativity. In the end, having found generally covariant field equations, he changed his view about the nature of the substantival elements of spacetime. To him, points of the manifolds cannot have an independent individuality. They can only be distinguished by reference to a field.

After he came up with his new perspective on the hole argument, Einstein did not worry much about the consequences. But this was not the end of the relevance of the hole argument. Its legacy led others to stumble as well. Earman, who used to be an advocate of substantivalism, collaborated with Norton; and in 1987 they claimed that the hole argument can be used to show that generally covariant field equations lead to radical indeterminism if one is a substantivalist. They argued that, because substantivalism is 'just' a philosophical doctrine, one should not give up determinism, which should only fail because of reasons of physics; but one should accept the principle of Leibniz equivalence and thereby deny substantivalism.

Butterfield responded to this threat to substantivalism and derived an alternative definition of determinism called Dm2. His definition is based on the initial value formulation of general relativity and is thus implicit in modern formulations of our best theory of spacetime. This new definition Dm2 reconciles substantivalism with determinism for generally covariant field equations; the hole argument is no longer a threat to substantivalism. Besides this technical success, Dm2 represents the intuitive idea of determinism as faithfully as does Dm1, which is a stronger definition of determinism in terms of models.

Butterfield's alternative definition Dm2 is not a makeshift, but rather an attractive, definition of determinism. Substantivalism takes the fundamental objects that a theory states to exist as indeed existing. General relativity quantifies over points of a spacetime manifold and therefore a substantivalist takes those points as genuine objects. Einstein's approach to take a combination of points and fields, that assign properties to the points, as fundamental is, I would suggest, unsatisfactory, because those fields have to obey the field equations of the theory. They are determined by partial differential equations and thus not an inherent part of the spacetime but are a physical field in it; i.e. the fields are a content of the spacetime. Dm2 allows a substantivalist to take 'only' points on the spacetime manifold as fundamental objects and still secures determinism from generally covariant field equations. This is, in my opinion, an attractive view of both substantivalism and determinism and should be endorsed by a scientific realist.

I would like to thank Daniel Toth for discussions, Farrah Raza for proofreading and Jeremy Butterfield for discussions and comments on previous versions.

References

[1] Anderson, James: (1967) "Principles of Relativity Physics", *Academic Press*, Chap. 4, 10

[2] Butterfield, Jeremy: (1987) "Substantivalism and Determinism", *International Studies in the Philosophy of Science*, **2**, 10-32

[3] Butterfield, Jeremy: (1989) "The Hole Truth", *British Journal for the philosophy of science*, **40**, pp.1-28

[4] Earman, John: (1977) "Leibnizian Space-Times and Leibnizian Algebras", in Butts and Hintikka (eds.): "Historical and Philosophical Dimensions of Logic Methodology and Philosophy of Science", *Reidel*

[5] Earman, John and Norton, John: (1987) "What Price Spacetime Substantivalism", *British Journal for the Philosophy of Science*, **38**, 515-525

[6] Einstein, Albert: (1961) "Relativity, the special and general theory", authorised translation by Robert W. Lawson, *Wings Books*

[7] Lewis, David: (1986) "On the plurality of worlds", Blackwell

[8] Norton, John: (1984) "How Einstein found his Field Equations: 1912- 1915", *Historical Studies in the Physical Sciences*, **14**, 253-316; reprinted in Don Howard and John Stachel (eds.): (1989) "Einstein and the History of General Relativity: Einstein Studies, Vol. 1" *Boston: Birkhuser*, 101-159

[9] Norton, John: (2001) "General Covariance, Gauge Theories and the Kretschmann Objection", *Preprint*, URL = <http://philsci-archive.pitt.edu/id/eprint/380> (4.4.2011)

[10] Norton, John: (Winter 2008 Edition) "The Hole Argument", *The Stanford Encyclopedia of Philosophy*, URL = <http://plato.stanford.edu/archives/win2008/entries/spacetime-holearg/> (20.2.2011)

[11] Wald, Robert: (1984) "General Relativity", *University of Chicago Press*, Chap. 10