DRAFT

There are no events without spaces and times: What Einstein learned from his "hole argument"

John D. Norton
Department of History and Philosophy of Science
University of Pittsburgh
https://sites.pitt.edu/~jdnorton/jdnorton.html

Standfirst: *What counts as real in our most fundamental physical theories? Professor of the history and philosophy of science at the University of Pittsburgh, John D. Norton recounts Einstein's struggles with his "hole argument" of 1913. Einstein first thought that the argument would compel him to abandon the most distinctive feature of his then incomplete general theory of relativity, its general covariance. He escaped his error by reaffirming a powerful method for distinguishing physical reality from mathematical redundancy. The elements in a theory that correspond to real things in the world are only those that remain unchanged when we alter our mathematical descriptions. Norton argues that we now interpret the significance of the hole argument as establishing that the events of spacetime do not form what philosophers define as a substance, that is, something that can exist independently of other things in the world. We need also to specify the times and distances between these events; and only the resulting totality forms a spacetime of our physical world.*

What is real? What is not real? These simple questions have long exercised philosophers of science. Are there quarks? Are there black holes at the centers of galaxies? Really? If they weren't real, wouldn't the enduring successes of their sciences be something of a miracle? We have played this game before. How could there be no Absolute Space, no ether, no phlogiston, and no caloric to make sense of how heat flows? We were certain about them all, but then we learned that we were wrong. Perhaps we are wrong again. Debates over what is real at the most general level of scientific theories do still endure.

When differences manifest in nothing observable, that fact has long been taken as a strong indication that the differences correspond to nothing real. In an intriguing episode in the history science, a new guide emerged. The theory itself can sometimes help us decide what is real in its domain. In 1913, Einstein used his "hole argument" to justify his then mistaken formulation of what would become his greatest achievement, the general theory of relativity. Einstein's original analysis was mistaken since he attributed reality to differences that were invisible even to his theory. Given the fullest surrounding information, his emerging general theory of relativity could not decide which of several possibilities were to be distinguished as the real one. Einstein ultimately concluded that his theory could not distinguish which was real because both corresponded to the same reality. Thus, both have to be equally real. His analysis provided a new and powerful way for us to determine when theoretical differences correspond to nothing real.

<u>The familiar cases: identifying what is real in our mathematical theories</u>

Perhaps the most familiar example in theories of space of differences that correspond to nothing real concerns our choice of coordinate systems. When we lay out a Cartesian coordinate system in a Euclidean space, we identify the points of the space with suitable values of the coordinates x, y and z. There is a special point in this coordinate system: the "origin," where x, y and z fall to zero. If we are too literal in our reading of these coordinates, we might say that this point is special. It has a preferred reality that distinguishes it from every other point in the space. It might be, we imagine, that space has a central point that is distinguished uniquely by this zero of the coordinates. It is not just an arbitrary label.

Of course, no one should think that this origin point is in any way different from all the other points in a Euclidean space. A formal manipulation is one way to secure this conclusion. A simple transformation adds constant values to each of the x, y and z coordinates and has the effect of moving – shifting or translating—the origin to a different point in the space. We can carry out all our geometric analyses just as well in this new coordinate system using precisely the same constructions. The shift of the origin point has made no difference to the geometric facts that matter: the distances and angles in geometric figures are unaffected. In Figure 1, the 3-4-5 right angled triangle keeps the same lengths of its sides and its right angle, no matter where we place the origin of the coordinate system. Any special reality attributable to the first origin point must then also be attributed to every other point, since all of them can become the origin of a translated coordinate system. The difference in the location of the origin is a difference that corresponds to nothing real.

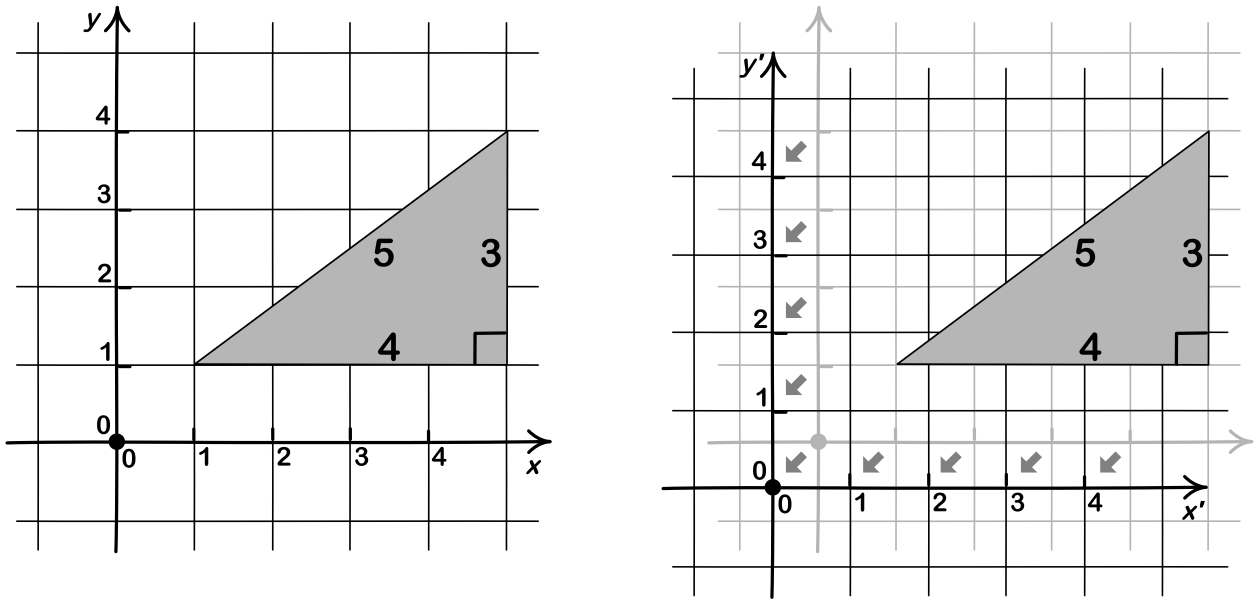

Figure 1. Two Cartesian coordinate systems with different origins

That shifting the coordinate system made no difference to the important geometric facts of matter, was not some *a priori* fact that somehow comes solely from the mathematics. It is a contingent matter that depends on the physical facts of Euclidean geometry: its space is homogeneous. No one point more 'special' than another. Things *can* go differently. Sometimes one point in a space *does* have some special physical status. The lines of latitude and longitude on the surface of the earth form a coordinate system that has two origin points. One is at the North pole and the other is at the South pole. These origin points are factually unique. They locate a physical feature of the Earth, the axis of the earth's rotation. We could shift the lines of latitude and longitude so that their origin points would be located elsewhere, as indicated in Figure 2. That shift would not change the nature of the two (original) origin points at the North and South poles. The facts of a spinning earth ensure the privileged status of these two polar points. They remain just where they have always been. In this case, there *is* a physically privileged status for some origin points over others.

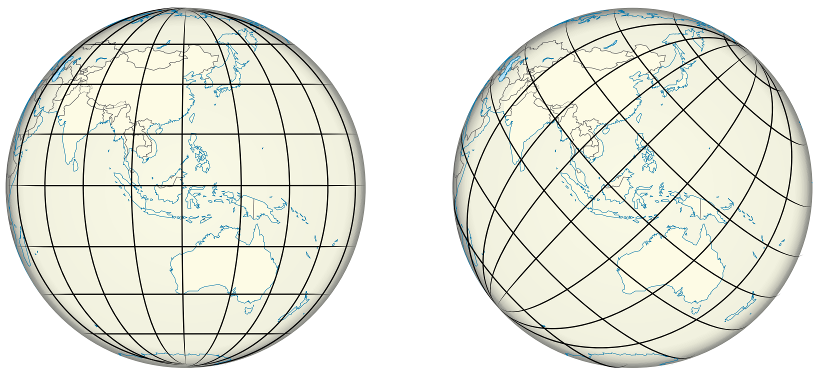

Figure 2. Original and displaced lines of latitude and longitude

Source image for globe.

https://en.wikipedia.org/wiki/File:Taiwan_on_the_globe_%28Southeast_Asia_centered%29.svg

The Hole Argument: Einstein's struggle with the realities of general relativity

Now to Einstein. These sorts of considerations played a major role in his discovery of the general theory of relativity. In the years leading up to the theory's completion in November 1915, he faced some difficulty in determining just what was real in his theory. The locus of his conern was what he called his "hole argument" [*Loch Betrachtung*]. A large modern literature has grown from Einstein's early deliberations and his resolution of the initial mistake.

To recount Einstein's struggles, it is most convenient to use a modernized formulation of his general theory of relativity. The theory portrays spacetime as having two elements. The first is a four-dimensional manifold of events. They are the individual points-events of spacetime, organized into continuous, nestled neighborhoods, as shown in Figure 3. Each event of the spacetime manifold is designated by one set of the values of the four-dimensional spacetime coordinate system. If such a set is the four real numbers <*t*, *x*, *y*, *z*>, then a neighborhood of surrounding points are just those events whose coordinates differ only slightly from the initial point's coordinates. Greater differences correspond to larger neighborhoods.

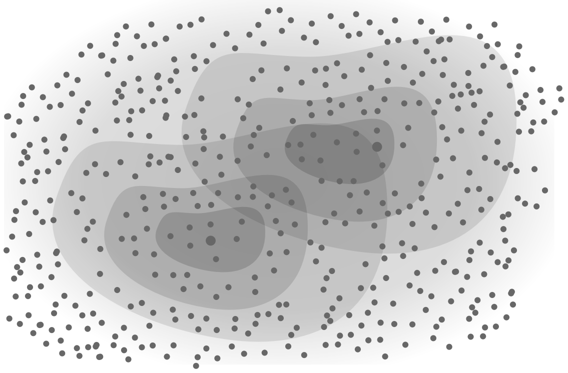

Figure 3. A manifold of spacetime events

The second element is the metrical structure. A manifold by itself does not tell us the distance in space or time elapsed between neighboring events. It does not even tell us which events are to the past or future of other events or happening at roughly the same time. The manifold is just a huge collection of points organized into neighborhoods. There is no space or time yet. We might image that we can use the differences in the coordinate values to tell us these spatial distances and times elapsed. It does not work. In formulating general relativity, Einstein used the newer mathematics pioneered in the nineteenth century by Gauss for his theory of curved surfaces. In that mathematics, all manner of coordinate systems could be used. The coordinate differences between neighboring events can take on a huge range of different values according to the coordinate system chosen. That range is far greater than the few numbers that could specify the distances in space and times elapsed between events.

The solution is to *add* the information explicitly in a new mathematical structure that gives, in compressed form, all the distances in space and time between neighboring events, as shown in Figure 4.

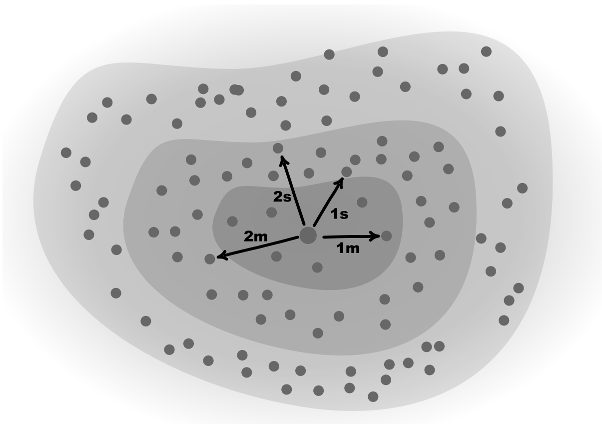

Figure 4. Manifold with the metric added for one point-event

This new structure is called the “metric” or “metric field” or “metric tensor.” If we know these distances and times for neighboring events, we can add them up as we move from event to event along some path. The addition of these many small quantities along the path will tell us the spatial distances between or times elapsed for any pair of events connected by the path we took.

The structure is called “*metric*” since the information it gives concerns the spaces and times that can be *measured* by ordinary measuring instruments such as rods and clocks. The rods measure so-called “proper distance,” which is the physical distance between events not separated in time. The clocks measure “proper time,” which is the time elapsed as a clock moves in spacetime between past and future events.

The main elements of the general theory of relativity are a spacetime manifold, the metric field and any additional matter fields defined on the manifold, such as would represent ordinary matter and energy in the spacetime. The relationship between these fields is given by Einstein’s celebrated gravitational field equations. They tell us how the presence of matter and energy causes spacetime to curve and they determine the ways in which this metric might be spread over the manifold of events, so that it conforms with the matter distribution in spacetime.

The distinctive feature, historically, of Einstein’s 1915 theory was its “general covariance.” Einstein followed Gauss’ liberalization in the choice of coordinate system usable by the theory. The expanded set allowed for all sorts of twists and distortions, even just those that applied only locally, as we shall see below. As we move between the descriptions of some physical spacetime provided by different coordinate system, many things will change in the descriptions. Since there is just one physical spacetime being described, real elements in the spacetime are just those aspects of the descriptions that do not change when we change coordinate systems. These unchanging elements are the so-called invariants of the coordinate transformations. The most familiar examples are spatial distances between events, as measure by rods, and the time elapsed between events as measured by clocks. These have to be invariants since they are measured by physical devices, independently of our choice of the coordinate system. Their invariance is akin to constancy under changes of Cartesian coodinates of the 3-4-5 triangle in Euclidean space discussed earlier.

The general covariance of Einstein's new theory opened up new possibilities. They were the very ones that led Einstein astray, initially. If we have a description of the metric and other fields in one coordinate system, we can create very many more descriptions of same physical spacetime just by transformations to different coordinate systems. The transformations that produce these new descriptions include the sort of global transformations that we applied to Cartesian coordinate systems in Euclidean geometry when we shifted the entire system as a whole. However, general covariance allows for a far richer set of transformations; some transformations might only shift the metrical structure locally, that is, in some small region of the spacetime.

A metaphor gives a useful mental picture of the scope of the transformations admitted by general covariance. We imagine that the point-events of the spacetime manifold are those of a tabletop; and that the metric is like a tablecloth spread over its surface. A global transformation would just shift the cloth as a whole. A local transformation might arise when we just distort or twist the tablecloth in some small area of table's surface, while leaving the rest of the tablecloth unaffected, as shown in Figure 5.

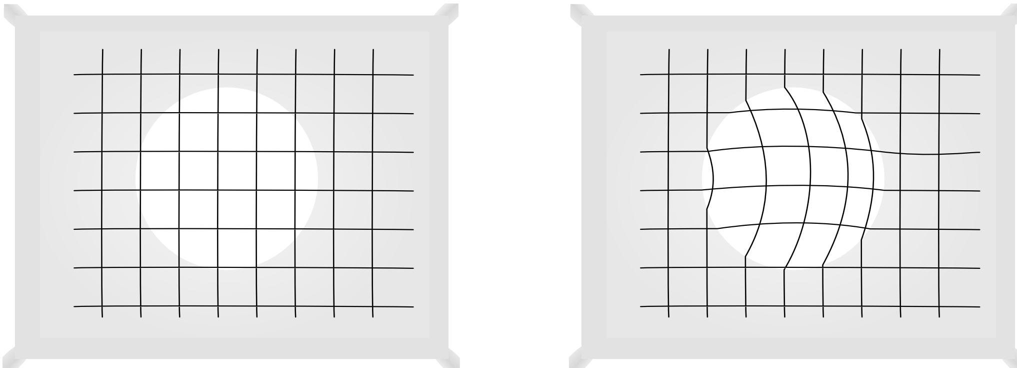

Figure 5. A hole transformation

Einstein considered just such a transformation in 1913. The small region in which the distortion occurred was taken to be a matter-free hole in the matter distributed over spacetime. The term "hole" was literal in his construction. With a few added steps, it was a simple matter for Einstein to use this transformation to realize what appeared to be two different metric fields within the hole. We can see the difference in Figure 6.

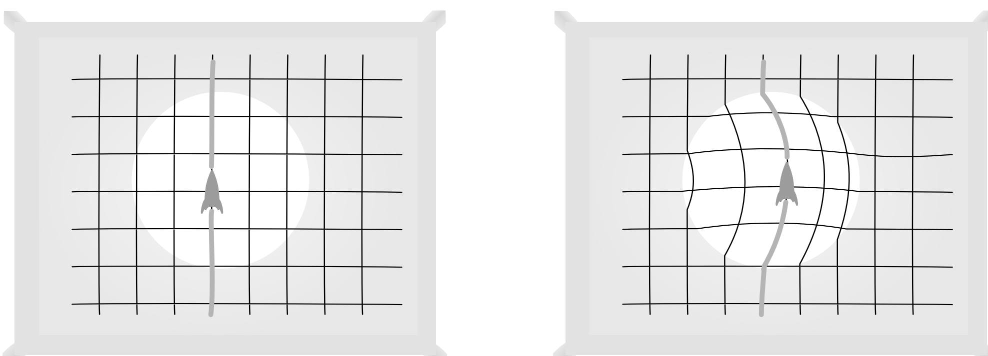

Figure 6. A spaceship traverses the hole.

It shows the worldline of a spaceship that passes through the hole; and the same spaceship after the transformation. In the two cases, the spaceship passes through different events in the part of the spacetime manifold within the hole. That difference appeared to Einstein to be a real physical difference. He knew that the two cases were related by a simple transformation, but he could not shake the idea that there was still a difference physically between the two cases.

All this played out in publications of 1913 and 1914, when Einstein was writing initially in collaboration with his mathematician-friend Marcel Grossmann. Together, they published the first sketch of what would soon be his final general theory of relativity. Pretty much all of the 1915 theory was already in the 1913 "*Entwurf*" ("Sketch"). It lacked one key element. Einstein and Grossmann had considered generally covariant gravitational field equations but they had rejected them on the mistaken grounds that they were incompatible with Newtonian gravitation theory in the limiting case of very weak, static gravitational fields.

Einstein was now in something of a quandary. The idea of generally covariant gravitational field equations remained appealing, but they were beyond his reach. He had initially failed to find what he deemed serviceable generally covariant gravitational field equations. Might it just be his failure to look hard enough for them? No! Einstein soon convinced himself that all generally covariant gravitational field equations are physically inadmissible. His theory should not use them. It is a conclusion astonishing to modern readers for whom general covariance is the signal achievement of Einstein's final theory!

The support for this now startling conclusion came from his "hole argument." The construction above depends upon the general covariance of the gravitational field equations. It is possible not just in the generally covariant theory of 1915 but in *any* generally covariant theory. It appears to show us something very general and very powerful. The striking thing, Einstein felt, about the construction sketched above is that there are factual differences within the hole. The spaceship (in my modern dramatization) visits different events in the two cases. Those differences, however, are quite localized. The two cases do not differ at all outside the hole.

This, Einstein concluded, was a failure of determinism. Given the fullest specification of the everything outside the hole, in its past, future and elsewhere, generally covariant gravitational field equations would fail to specify physical facts within the hole. In 1913, Einstein conceived determinism as the same as causality. He could then lament that generally covariant gravitational field equations lead to violations of causality. Therefore, he concluded general covariance was not a physical possibility.

Einstein's escape

Einstein's *Entwurf* theory was by modern lights a malformed theory and, at a visceral level, Einstein knew it. As the months and then years passed, he continued to try to convince himself that all was well with this flawed theory. By November 1915, the failures of the theory had so mounted that Einstein took a major step. He abandoned his gravitational field equations of 1913 and, in four, rapid-fire communications to the Prussian Academy in that November, arrived at the generally covariant gravitational field equations of the modern theory. In the third of these communications, Einstein reported the greatest moment of his scientific life: he found that his new theory now accounted exactly for the anomalous motion of the planet Mercury.

While he celebrated this great triumph, a now rueful Einstein had the awkward job of explaining to his correspondents and readers what went wrong with his earlier analysis of the hole argument. His solution was that he had mistaken the realities within the hole. When his transformation reassigned metrical properties to events in the hole, the difference was a difference that made no difference physically. There was nothing in any possible observations that could pick apart the original and transformation solutions. There was no reality in the theory that distinguished points of manifold beyond the spatio-temporal metrical properties assigned to them by the metric.

In this sense, Einstein's transformation was akin to the shifting of the Cartesian coordinate system in the example above. Einstein, however, had mistakenly treated it as we would the shifting of the lines of latitude and longitude over the surface of the earth. The particular lines of latitude and longitude are chosen because their origins reflect the physically meaningful location of the Earth's North and South poles. Shifting these origins undoes this important physical connection. In this aspect, they are unlike the origin of the Cartesian coordinate system.

The transformation between the two descriptions in the hole argument corresponds to no real changes in the physical spacetime. It may appear that the spaceship passes through different spacetime events in the two cases. But that is illusory. We cannot associate a point in the mathematical manifold with a real event in the physical spacetime independently of the metrical properties present at that event. When we transform between the two cases, we also move the metrical properties assigned to the points in the manifold. As a result, the spaceship passes through the same physical events in the two cases, even though it appears otherwise in a simple reading of the two descriptions. In short, all physical facts depicted by mathematical structures within the hole are same. There is no failure of determinism since there are no physical facts left underdetermined within the hole

What we learned

Einstein's struggles with his hole argument has left us with an important legacy that has proven fertile still today. It happens quite often that a physical theory can give us two closely related solutions of its fundamental equations. How are we to know whether they represent the same reality, just described differently, or whether they represent two distinct realities? The decision between the two cases has to be made on physical grounds. The case of general relativity is typical. It is a physical question not one of simple mathematics. It must be decided by physical criteria.

Einstein's analysis gave us two criteria for the decision.

The first had long been recognized. Whereas the two solutions may be different mathematically, is there anything observational that distinguishes them? If not, we have our first basis for concluding that the mathematical differences do not correspond to any differences in reality. This criterion lay behind the enduring awkwardness of Newton's conception of Absolute Rest. Nothing observable could pick out that state of rest from many possible inertial motions.

The second criterion was new and powerful. In the case of the hole argument, it was not just that we could observe no difference, that is, that nothing observable distinguished to two cases. The theory itself was unable find enough of a difference for it to be able to pin down any supposed reality that might separate the two cases. For as we saw above, even when given the fullest specification of everything outside the hole, Einstein's theory could not decide which of the possible extensions into the hole was the right one.

Einstein's problem in discerning what is real is an instance of a familiar problem in modern physics. Under the guise of "gauge transformations," we are able to produce multiple solutions of a theory that differ in ways analogous to those Einstein found. Do they reflect a difference in what is real? To answer we can now ask: "Does anything observable separate the two cases? Can relations of determination with the theory privilege one of them?" If the answer to both is no, then we have a strong basis for concluding that nothing real separates the two cases.

The outcome of Einstein's deliberations on the hole argument also tells us something about the spacetime events of theories like general relativity. It is tempting to image that the events in the mathematical spacetime manifold correspond to specific physical events without further qualification. If this independence were the case, then problems ensue. A generally covariant theory like general relativity allows us to take the metrical properties that happen to belong to one spacetime event and reassignment to others. What would result, we now learn through the example of the hole argument, is that the theory becomes indeterministic. It can no longer determine which metrical properties belong to which events. The escape—Einstein's solution—is to accept that we cannot associate points in the mathematical spacetime manifold with real events in the physical spacetime without also considering the metrical properties associated with the event. That is, an essential part of the identity of a spacetime event resides in its spatial and temporal distances from other events. It is not the event it is without these spaces and time. There are no events without spaces and times.

Reading: Norton, John D., Oliver Pooley, and James Read, "The Hole Argument", *The Stanford Encyclopedia of Philosophy* (Summer 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/sum2023/entries/spacetime-holearg/>.

Norton, John D. "Ontology of Space and Time: The Hole Argument" in *Einstein for Everyone*. https://sites.pitt.edu/~jdnorton/teaching/HPS_0410/chapters/significance_GR_hole_argument/significance_GR_hole_argument.html