# The Delicacy of Counterfactuals in General Relativity (Or: If General Relativity Were Not Difficult, Its Counterfactuals Would Be Easy)

Erik Curiel†

October 31, 2014

> If wishes were horses, beggars would ride.
>
> Proverb

**ABSTRACT**

General relativity poses serious problems for counterfactual propositions peculiar to it as a physical theory, problems that have gone unremarked on in the physics and in the philosophy literature. Because these problems arise from the dynamical nature of spacetime geometry, they are shared by all schools of thought on how counterfactuals should be interpreted and understood. Given the role of counterfactuals in the characterization of, *inter alia*, many accounts of scientific laws, theory-confirmation and causation, general relativity once again presents us with idiosyncratic puzzles any attempt to analyze and understand the nature of scientific knowledge and of science itself must face.

In his elegant, magisterial exposition of the foundations of general relativity, Malament (2012, ch. 2, §1, pp. 120–121) provides three interpretive principles to endow the mathematical framework of Lorentzian geometry with physical content:[1]

> For all smooth curves $\gamma : I \to M$ [where $I \subset \mathbb{R}$ is an open interval and $M$ is a candidate spacetime manifold]:

† **Author's address**: Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität, Ludwigstraße 31, 80539 München, Deutschland; **email**: erik@strangebeautiful.com

[1] I follow Malament (2012) in all relevant conventions (the signature of the spacetime metric, the definition of the Weyl tensor, *etc.*). The reader should consult that work or Wald (1984) for exposition of all concepts and results about general relativity I rely on in this paper.

**(C1)** $\gamma$ is timelike iff $\gamma[I]$ could be the worldline of a point particle with positive mass;

**(C2)** $\gamma$ can be reparametrized so as to be a null geodesic iff $\gamma[I]$ could be the trajectory of a light ray;

**(P1)** $\gamma$ can be reparametrized so as to be a timelike geodesic iff $\gamma[I]$ could be the worldline of a *free* point particle with positive mass.

(Emphases are Malament's; 'C' indicates the proposition pertains to the interpretation of conformal structure, 'P' to projective structure.) He immediately offers four comments and qualifications to address possible concerns one may have with these propositions as interpretative principles, touching on questions about the exclusion of tachyons, the restriction to smooth curves, the status of point particles in general relativity, and, of most interest for our purposes, the modal character of the propositions. He concludes (*ibid.*, p. 122),

> Though these four concerns are important and raise interesting questions about the role of idealization and modality in the formulation of physical theory, they have little to do with relativity theory as such.

I agree with his conclusion in all parts, except for the concern about the role of modality. I think there are important problems with modality in general, and with the understanding of counterfactuals in particular, peculiar to general relativity as a physical theory, problems, moreover, that have gone unremarked in the philosophy and the physics literature. Malament's formulation and discussion of the interpretative principles allows them to be drawn out with great clarity.[2]

About (C2) he says (*loc. cit.*), "We are considering what trajectories are available to light rays when no intervening material media are present—*i.e.*, when we are dealing with light rays in vacuo." Now, surely we want to talk as well about the null cones even at those places where matter is present. In order to do so, and in order to formulate the analogue of (C2) for those spacetime regions (in order to give a physical interpretation to the null cones at those points), we must say something along the following lines: the null geodesics where matter is present are those paths light rays would follow if the matter there were removed. But on its face, that modal statement makes no sense in the context of general relativity, because however we make sense of the idea of "removing matter" from a spacetime region, the metric will *eo ipso* be different in that region from what it was, and it will generically be the case that the new metric in that region will not agree with the original metric on what it counts as null vectors, among many other differences.[3] The distribution of matter in a region of spacetime in large part informs the metrical structure there, so what sense can be made, in the context of the theory, in asking what the metrical structure *would* be if the matter actually there *were not* there? And now we face the heart of the problem: the ineliminable ambiguity in the

[2] I want to emphasize that I am not criticizing Malament or trying to draw attention to weaknesses or errors in his exposition, quite the contrary. It is the exemplary (and characteristic) clarity, precision and thoroughness of his discussion that allows a previously unacknowledged problem to be brought into the light.

[3] The same problem arises for timelike curves in regions of spacetime already occupied by matter, *i.e.*, for (C1) and (P1), but I will focus on the case of null rays to simplify the exposition.

idea of what it may mean to "remove matter from a region of spacetime" guarantees that we have no way to conclude on any principled basis "what the metric would then look like there".

The problem is made more acute by the fact that metrical curvature is *only* in part informed by the distribution of matter: the Weyl curvature at a point, exactly that part of the curvature encoding conformal information, such as what counts as a null vector, is independent of the value of the stress-energy tensor at that point—the value of the Weyl tensor, point by point, is not constrained by the presence or absence of matter. In regions without matter, moreover, metrical curvature is governed entirely by the Weyl tensor. Still, the Weyl tensor $C^a{}_{bcd}$ is subtlely related to the distribution of matter at neighboring points, when there is such matter, in a way that can be made precise by using the Bianchi identity formulated using the so-called Lanczos tensor.[4] Thus, in "removing matter" from a spacetime region, there can be no principled way to determine what the "remaining curvature" will be. One may decide to keep the Weyl tensor the same. But precisely its relation to stress-energy by way of the Lanczos tensor means that this is not an unproblematic way to proceed, and is likely even incoherent or inconsistent.[5]

To make the problem a little more concrete, consider the following situation: one has a solid translucent medium of a fixed shape and material constitution, and one wants to know "the path a light ray would take if one drilled a narrow cylindrical hole through the middle and pointed a laser through it". There are (at least) two obvious ways of preceding. First, one can look for a static solution for such an object (*i.e.*, a solution for such an object "existing in isolation for all time"), and then solve, for some fixed interval of "time", for the null geodesics that thread the 4-cylinder formed by the opening. Second, one can model the process of drilling the hole in the initially solid object, and similarly solve for the relevant null geodesics (at a region far enough to the future of the drilling process to ensure that all gravitational excitations induced by the drilling have had time to radiate away, *etc.*). In general, there is no reason why the two modeling procedures will yield the same answers to any questions one may want to pose about the null geodesics threading the opening. (One must be careful here: since one is dealing with two different spacetimes, one cannot simply "compare the null geodesics yielded by each procedure", and, indeed, the kinds of question one will be able to ask so as to be able to sensibly compare answers the two procedures give is limited.) And there is no principled reason to prefer one of the procedures to the other.

[4]The Lanczos tensor is defined as follows:

$$\begin{aligned} J_{abc} &:= \frac{1}{2}\nabla_{[b}R_{a]c} + \frac{1}{6}g_{c[a}\nabla_{b]}R \\ &= 4\pi\nabla_{[b}T_{a]c} - \frac{1}{12}g_{c[b}\nabla_{a]}T \end{aligned} \tag{0.1}$$

then the Bianchi identity may be rewritten

$$\nabla_n C^n{}_{abc} = J_{abc}$$

Thus the value of the Weyl tensor at a point does depend in an indirect way on the distribution of matter at nearby points.

[5]It should therefore be clear that these sorts of problem arise not only for counterfactuals involving changes in the distribution of matter, but also for any involving changes in the curvature more generally. One may, for example, try to consider how the behavior of test-particles in a vacuum spacetime would change if one were to "remove a component of the ambient gravitational radiation".

It follows that there is no single algorithm or reasoning procedure to employ for all such possible cases of proposed counterfactual reasoning in general relativity. One will have to handle such situations on a case-by-case basis, coming up with a method to fix the metrical structure in the proposed counterfactual situation in some way that respects the particular constraints of the project the counterfactual reasoning is to play a part in.

Compare the situation in Newtonian gravitational theory. It makes perfect sense in Newtonian theory to reason counterfactually about the behavior of a given kind of system in the presence or absence of any other kind of system, since that presence or absence won't affect the kinematical structure of Newtonian spacetime. There is, for example, no problem in principle in computing the counterfactual change in gravitational forces in a region induced by any counterfactual changes in the distribution of matter anywhere in the spacetime. But one just cannot do that in general relativity, unless one spells out what the new metrical structure will be in advance when one tries to reason counterfactually about what would happen if one were to "change the distribution of matter in a region of spacetime".

It is important to note, nevertheless, that there is at least one kind of counterfactual in the context of general relativity that does not face these sorts of problems: those that do not involve counterfactual changes to the distribution of matter or the structure of curvature more generally. As Malament himself notes (*ibid.*, p. 121, his emphases), "It is simply not true—take the case of (C1)—that all images of smooth, timelike curves *are*, in fact, the worldlines of massive particles. The claim is that, as least so far as the laws of relativity theory are concerned, they *could* be." One way to make such a modal claim precise is to invoke the notion of a test-particle, *i.e.*, by assuming that the particle is so small and has so little stress-energy that it is an excellent approximation to ignore any contribution it could make to the stress-energy tensor, leaving the metric unchanged along the path at issue.

The problem I expose in this paper is severe: many influential philosophical approaches to many fundamental problems and issues in the philosophy of science—the nature of scientific laws, of theory-confirmation, of causation, *et al.*— rely, in ineliminable ways, on subjunctive conditionals for their formulation and application. Physicists certainly rely on such propositions in theoretical and experimental practice to propose and perform tests of general relativity. What reason do we have to believe that we understand what is happening in such cases in the context of general relativity, much less to have confidence in any conclusions drawn? Indeed, I think the situation is even worse than the preceding remarks suggest. Because the problem arises solely from the dynamical nature of spacetime geometry in general relativity, what I say here is wholly independent of one's favorite account of counterfactuals—it depends only on the theoretical resources general relativity provides to model such situations and pose such propositions, no matter what ancillary tools or frameworks one uses to interpret and understand them.

I wanted in this paper only to draw attention to this serious problem, not to propose possible solutions. I think any decent attempt to do the latter will require a great deal of involved, technical work, including detailed examination of many non-trivial examples. I sincerely hope someone takes up the challenge. The mettle of philosophy and the needs of physics demand we understand what is going on here.

## References

Malament, D. (2012). *Topics in the Foundations of General Relativity and Newtonian Gravitational Theory*. Chicago: University of Chicago Press. Uncorrected final proofs for the book are available for download at http://strangebeautiful.com/other-texts/malament-founds-gr-ngt.pdf.

Wald, R. (1984). *General Relativity*. Chicago: University of Chicago Press.