The Constraints General Relativity Places on Physicalist Accounts of Causality†

Erik Curiel

November 1, 2011

Those who make causality one of the original *uralt* elements in the universe or one of the fundamental categories of thought,—of whom you will find that I am not one,—have one very awkward fact to explain away. It is that men's conceptions of a Cause are in different stages of scientific culture entirely different and inconsistent. The great principle of causation which we are told, it is absolutely impossible not to believe, has been one proposition at one period of history and an entirely disparate one another and is still a third one for the modern physicist. The only thing about it which has stood, to use my friend Carus' word, a κτῆμα ἐς ἀεί,—*semper eadem*—is the *name* of it.

Charles Sanders Peirce
Reasoning and the Logic of Things

Contents

†This paper, in a slightly unmitigated form, was published as Curiel (2000).

1 Introduction

It is well known, at least among those who think about these things, that energy is not a fundamental quantity in the theory of general relativity. By the claim that energy is not a fundamental quantity in general relativity, I mean that within the mathematical structure of the theory one cannot rigorously define a quantity that has any of the features one might take to be definitive of energy. This does not imply that one cannot talk about energy at all within the theory. The startling fact, however, is that only in certain special situations can one represent within the theory a quantity that is structurally similar to energy as it is manifested in classical physics[1] and special relativity, and even then only by employing explicitly approximative and idealizing techniques that are not part of the theory *per se*—the Einstein field equation does not of its own nature exhibit simplifying modifications to itself appropriate for the weak gravitational field regime, much less justify the use of such modifications in any particular circumstances. Consequently talk about energy reflects nothing fundamental about the theory itself.

In this paper, I discuss how this fact may place restrictions on the sorts of accounts of causality that can be considered fundamental *vis-à-vis* general relativity. Some accounts of causality purport to treat only relations among middle-sized dry goods in everyday practical affairs; others take causality to be something akin to a logical category of thought that structures our knowledge of various matters; yet others take 'causality' merely to indicate that a special type of explanation is required or is in the offing; and yet others take it as a merely subjective, psychological phenomenon, the manifestation of a brute fact about the way we are constructed to view the world. The arguments of this paper do not pretend to bear on any such accounts. Certain sorts of accounts of causality, perhaps best exemplified by Russell (1927), rest on the idea that causality is a *physical* relation holding among *physical* entities, and as such must accord with best going physical theory. It is only such accounts that concern me here. Many such accounts rely on the intuitively appealing idea that energy and other classically conserved physical quantities such as momentum are intimately connected with causality, in so far as this concept represents actual structure of the physical world as modelled by our best physical theories. Especially popular is the idea that the propagation and transfer of energy, and all energetic processes in general, *embody* certain sorts of causal relations. For the sake of brevity I will refer to all such accounts of causality as *transfer accounts*, gesturing at the fact that these sorts of accounts take the transfer (and the propagation) of energy to play the most important role in constituting the causal relation. In so far as such accounts hold this and similar tenets, I argue, and in so far as one considers general relativity, or at least certain aspects of it, to constitute fundamental physical theory, transfer accounts cannot be correct.

My arguments, though likely affronting to some dearly held contemporary intuitions, should not, I think, be terribly surprising. The ideas of propagation and transfer, as we understand them, have not been associated with the notion of causality commonly held by the intelligentsia for long,

[1] When I say that a physical theory or system is 'classical', I mean that it finds its proper representation in the spacetime of pre-relativity physics and that it does not fall under the purview of quantum mechanics.

certainly not for more than 400 years, since the time of Galileo.[2] Before the scientific revolution, such ideas were no part of generally held conceptions of causality among those who contemplated such matters. It was only in response to the development of classical mechanics that these ideas began to ingratiate themselves widely, and to acquire the honorific 'intuition', with its attendant privilege—future argumentation and theorizing had to conform now to these principles, but they themselves did not stoop to be questioned. The wide acceptance of "action-at-a-distance" pictures of gravitation and electromagnetism in the eighteenth and nineteenth centuries even suggests that these ideas were not regarded as intuitions until quite recently, perhaps beginning only with the general acceptance of Maxwell's electromagnetic theory, after Hertz's landmark experiments in the 1870's proved the existence of electromagnetic radiation.[3] Such intuitions would then have hit their stride with the acceptance of relativity theory in the early decades of this century, the more troublesome aspects of quantum theory being conveniently overlooked.

I do not think we should feel any reluctance to jettison some or even all of the contemporary notions of and intuitions about causality, in so far as they may apply at a fundamental level. At least, we should feel less reluctance to give up lessons learned early and often enough by the child to be termed 'intuitions' by the adult, than we should feel to ignore what our best physical theories seem to be trying to tell us about the character of the world. Philosophers, when having investigated questions of this sort, have tended to focus their attention on the lessons to be drawn from quantum mechanics. There are natural reasons for this, perhaps the most important being the great difficulty of the technical machinery one must master in order to study general relativity in any depth and with any breadth, as compared to that required for the examination of (non-relativistic) quantum mechanics. When they do turn their attention to general relativity, it too often happens that philosophers approach it with a raft of traditional concepts, questions and issues already in hand and ask how general relativity bears on them, as in the case of the time-worn debate on substantivalism versus relationalism, rather than studying general relativity both broadly and in detail to see what questions the theory itself naturally suggests we attend to and what concepts and structure it offers up as the most natural in the terms of which to formulate these questions. Many traditional questions and issues do not seem to get even so much as a secure footing in general relativity when the time and care is taken to ascertain what sort of structures one needs to have in place to frame them sensibly; and many questions longing to be asked remain wall-flowers.[4]

This paper occupies a funny middle ground between what I have painted as virtue and vice: it asks no wall-flowers to dance, and it does focus on questions and concepts originally mooted in the

[2] I do not think that the medieval doctrine of the identity of cause and effect involves a notion of continuity even vaguely analogous to ours. The character of the concepts in the terms of which medieval philosophers formulated their ideas and arguments differ so much from those used today as to make meaningful comparison impossible.

[3] I fault Hume's misunderstanding of Newtonian theory for blinding much of the philosophical world to this fact, even down to this day. See Stein (1994) for a brief discussion.

[4] Many of Russell's remarks in *The Analysis of Matter* (1927) and Eddington's in his *Mathematical Theory of Relativity* (1921) show a great sensitivity to general relativity's demand for new questions and new issues, and not a stale re-hashing of centuries old debate. Many contemporary philosophers, including Stein (npub), Malament (1977), Earman (1995) and Norton (1985), have what I think are similar views. I believe, among contemporary physicists, Penrose (1968) and Geroch (1973) do as well.

theater of classical physics; it does nevertheless attempt to show why this strategy can be misguided, by arguing that the classical concepts used to frame the questions do not fit within the structure of general relativity without much shoe-horning and discomfort. This paper does not pretend to be a mathematical demonstration of a certain philosophical point. Indeed, given the nature of the point I am arguing for, I do not think such a demonstration could be had either for its truth or its falsity. For the point is about the way that a certain picture of classical physics and of special relativity naturally suggests that causality has a certain character, whereas general relativity does not share the features of those theories that made the suggestion plausible.[5]

I begin in §2 by discussing the motivations behind transfer accounts of causality. In §3 I attempt to make precise in a schematic way the relations between energy and causality that transfer accounts presuppose. I will not attempt to give an exact analysis of energy as represented in any particular physical theory; rather I will assume some general propositions about energy that are part of the widely accepted folklore about physics, just to get the ball rolling. Next, in §4 I examine in detail the properties energy must have in order to play the roles demanded of it by transfer accounts. I will argue that, if the relations introduced in §3 are to hold between energy and causality, then one must be able to formulate something like the classical principle of the conservation of energy. A slight generalization of this argument will show that an appropriate conservation law is the necessary condition for any stuff's standing in such a relation to causality, from momentum to matter to electromagnetic field intensity—if one wants the relation of cause to effect to be mediated by the propagation and transfer of some stuff from the one to the other, then that stuff must be capable neither of being created nor of being destroyed. Finally, in §6 I turn to the theory of general relativity. In the theater of a generic general relativistic spacetime, not only can such relations not hold between causality and energy as transfer accounts require, but such relations cannot hold between causality and *any* stuff representable in the theory. The structure of general relativity, in so far as it precludes the requisite sorts of conservation principles, militates against this type of account of causal relations. I conclude with a brief description of the sorts of conceptions of causality that general relativity is not overtly hostile to.

Before beginning these arguments, I should say a few words about quantum mechanics, and why I will steadfastly ignore it after these few words. The aim of this paper is twofold. The first goal is to try to determine what sorts of accounts of causality may be consonant with our best going science. General relativity by itself cannot be considered a completely fundamental physical theory; it is hoped and expected that it will in due course be replaced by a theory that unifies gravity with the other three fundamental forces now so successfully treated by the Standard Model of quantum field

[5] Many among both philosophers and physicists have retrojected the idea of locality (widely accepted in the physics community only since the turn of the century) into all facets of classical physics, and have used the idea as a foundation in their analyses of causality even when they consider only classical phenomena. Since this intuition lies behind the sorts of accounts of causality I critique, but mostly because it makes classical theory more similar to relativistic theory, which we have good reason to think expresses a deeper understanding of the physical world, I will treat classical systems as interacting locally, and classical physics in general as conforming to a principle of local action. Of course, this requires that I bracket Newtonian gravitational theory entirely, since it is a non-local theory down to the bone.

theory. Although we have little if any hard evidence indicating any specific attributes such a deeper theory ought to manifest—the grandiquolent claims of string theorists notwithstanding—there is good reason to think that it will mirror certain fundamental features of both general relativity and quantum field theory that are too experimentally entrenched to envision discarding, just as general relativity itself mirrors certain fundamental aspects of Newtonian gravitational theory, *e.g.*, the roughly $\frac{1}{r^2}$ dependence of the acceleration of mutually gravitating, proximate bodies in the weak gravitational field regime. I cannot say with certainty that the absence of conservation principles of a certain sort is that kind of fundamental feature of general relativity which must find analogous expression in whatever better theory comes along to subsume it. Given that this feature of general relativity follows directly upon the presence of curvature (in a certain technical sense) in the metrical structure of spacetime, in conjunction with the fact that the primary energetic quantity of the theory is a two-index tensor and not a scalar—as fundamental a pair of features of the theory as one can well imagine—it does not seem to me a foolhardy bet. Even if in the event it turns out not to be so and principles appear in that more fundamental theory capable of supporting transfer accounts of causality, I do not think my work here will have been a waste.

This brings me to the second goal of the paper: to acquire a better understanding of general relativity itself as a physical theory. It is perhaps not widely appreciated how poorly understood general relativity is in many respects *vis-à-vis* quantum mechanics. The greatest obstacle, as already remarked, is the more difficult mathematics involved in mastering the theory and employing it to make physical predictions and retrodictions we have strong reason to believe are accurate. That, in conjunction with the poor experimental access we have to phenomena in strong gravitational fields, means that we have no good way to adjudicate among a swarm of competing schemes for, *e.g.*, providing approximate solutions to the general relativistic equations of motion for small, dense bodies; producing closed, analytical solutions to these equations is out of the question.[6] We also have only a superficial understanding of how to catalogue solutions to the Einstein field equation according to physically relevant, generic features they may share.[7] I think that neither the arguments nor conclusions of this paper constitute a real advance in our comprehension of general relativity, along these or any other lines. I rather hope only that it may make a small part of what we do comprehend available to a broader audience.

These difficulties by themselves would perhaps not suffice as a reason to ignore the tenets of quantum mechanics in so far as they bear on accounts of causality, but I think they do when taken together with the character of what it is we do know about quantum mechanics: in a direction orthogonal to the considerations of the previous paragraph, quantum mechanics is far less well understood than general relativity. The so-called measurement problem appears to demand an 'interpretation' of quantum mechanics in a way not required for an understanding of how to model phenomena in general relativity. Foundational questions in quantum mechanics at the moment are

[6]See *e.g.*, Quinn and Wald (1999) for a survey of the issue. To be more precise, the problem is usually posed as finding solutions to appropriately simplified (*i.e.* 'idealized'—a nobler term) equations of motion.

[7]See Curiel (1999) for a brief discussion of this point, in the context of examining proposed classifications of singulur structure in relativistic spacetimes.

so turbid and disputed as to allow a defense of almost no clear, precise propositions about causality, except for the obvious: proceed with extreme caution. Perhaps this paper may make the case by example that this need not be the case in general relativity.

2 Causality and Energy

The 19th century discovery of the principle of the conservation of energy not only had immediate and wide-ranging ramifications into most areas of physics as known at the time, as Helmholtz (1853, *passim*) himself, one of its discoverers, was at pains to stress from the beginning; it was also recognized that it ought to bear on a proper philosophical analysis of causality.[8] Mill, for instance, in the eighth edition of his *Logic*, took the opportunity to add a section to discuss the bearing of the principle on his own account of causality.[9] The considered opinion of several respected philosophers of this century has also been that there exists a close connection between causality and energy, even when they have not promulgated accounts of the simple form "A is the cause of B if and only if A and B stand in such-and-such a relation as regards energy". They have tended to share the view that most, if not all, causal relations involve in some essential fashion a transfer of energy between cause and effect, and that this physical fact underlies much of the explanatory force of causal laws. Quine (1973, esp. p. 5), for instance, believes that the imparting of energy is the central idea in our common "causal idiom," and that the flow of energy provides a "root notion" of causality itself. Salmon (1984, esp. p. 146), inspired by Reichenbach (1956), comes close to identifying causal processes as exactly those that transmit energy. Russell (1927, 1948) and Reichenbach (1956) predicate their conclusion that transfer of energy is intimately connected to causal relations upon a thorough analysis of scientific theory, including the working through in detail of examples of energy propagation and transfer in various physical situations.

On the face of it, the idea that energy and causality, as these notions may arise in or be suggested by physical theory, are somehow intimately related to each other has much to commend itself. From a naive viewpoint, the natural measure of the quantity of available energy stored in a dynamical system—how much work it can do on other dynamical systems (ignoring the constraints imposed by the second law of thermodynamics)—by itself suggests such a relation: a cause is, roughly speaking, something that produces a change in something else, and changes require work. Of more philosophical interest, the fact that energy appears to propagate and be exchanged continuously holds out the hope of offering an explanation of one of the most deeply rooted and widely held beliefs about the character of causality, that it itself is manifested continuously: that between any

[8] It is a suggestive—and highly obscure—fact that one of the other primary developers of the principle of the conservation of energy, Mayer (1842, *passim*), relied heavily on arguments based on the scholastic doctrine of the continuity and indestructibility of causes, as encapsulated in the apothegm *causa æquat effectum* ("causes equal effects").

[9] Though Mill refers to it as the 'Principle of the Conservation of Force'; *cf.* Mill (1874, preface, p. viii). He concludes in the added section (book III, chapter V, §10) that in fact nothing requires alteration in his original account of causality *per se*, though it does afford the possibility of an interesting elaboration of it, *viz.* by providing a criterion to winnow in certain situations true causal chains from spurious correlations.

entity A and any other entity B such that A causes B there is always a third entity Q such that A causes Q and Q causes B.[10] A related principle holds that, if A causes B, where A and B are spatiotemporally separated, then the causal efficacy "travelled" from A to B via Q and *not* via Z, where Q is "spatiotemporally between" A and B and Z is not. The thought is that one can draw a more or less narrow tube through spacetime both such that the loci of all links in the causal chain lie inside it, and such that it respects the null-cone structure of spacetime ("the causal efficacy propagates no faster than light"). All propositions of this sort I will group together under the rubric *the principle of causal continuity*. I am convinced that most of the attraction of transfer accounts of causality arises from this idea of causal continuity and its "explanation" by reference to the continuity of energetic processes and their respecting of null-cone structure.[11] Last, but certainly not least, the linking of energy and causality would appear to provide a straightforward solution to one of the most vexed problems in this vexed area of inquiry, that of winnowing out the true causal processes and chains from sequences of correlated junk—the true causal chains are those along which energy propagates in the proper fashion. This alone would make it worth fighting for.

I think Russell and Reichenbach were right to try to ground such views of causality upon a thorough analysis of fundamental physical theory, including the working through of several concrete examples, for such views presuppose several substantive theses about the structure of the physical world (or at least about the structure of our best physical theories): that, for example, a dynamical quantity called 'energy' exists and has at least some of the properties we naively associate with it; in particular, that packets of it can be transferred between dynamical systems in such a way that one can not only keep track of the identity of certain packets of it for at least brief periods of time, but also so that one can determine, in at least a certain type of case, what was the source and what the sink of a given bit; that particular segments of the time-evolution of dynamical systems can be distinguished as those during which the system is 'isolated from its environment'; that when a system is 'interacting with its environment' it is possible, in at least some cases, to identify the precise bits of the environment it is interacting with, which is to say the source or sink of any energy it is acquiring or losing during the interaction; and so on. None of these seem on their face particularly contentious. A conception of causality, however, that requires such theses and that has the stated goal of being fundamental—in the sense that it finds its motivation or justification in fundamental

[10] In fact, this statement says only that causal chains are *dense*, not necessarily continuous. A precise statement of the necessary and sufficient conditions for the continuity of causal chains would involve elaborate and unnecessary technicalities. I trust the sense of what I intend is clear without them.

[11] It is astonishing how widespread such an assumption of causal continuity is in the philosophical literature, especially when one considers how recently physics adopted even remotely analogous principles. Under the influence of Newton's theory of gravity, action-at-a-distance theories reigned until the late 19th century, and even today such comprehension of quantum mechanics as we have does nothing to encourage such views. Philosophers as disparate in temperament and aim as Ducasse (1926), Russell (1927), Russell (1948), Reichenbach (1956), Quine (1973), Mackie (1980), Bunge (1979), Salmon (1984), Lewis (1986), Mellor (1995) and many others too numerous to mention have all invoked in discussions of causality, more and less crucially, a principle of causal continuity. Perhaps most striking of all is that, among this whole lot, only Russell, in both works, discusses the possible grounds for holding such a principle and the consequences of its falsity, if it should turn out so. Every other philosopher takes it as an *a priori* principle from which conclusions about causality are to be drawn, but which itself need not, perhaps cannot, be questioned.

physical theory—just because it presupposes such substantive theses about the physical character of the world, requires for its justification a thorough investigation of the best going current physical theory or theories, if not to ground it satisfactorily at least to demonstrate that such assumptions as these, and whatever others the analysis of such an account of causality turns up, do not overtly conflict with any of the precepts of our best current science. One would for instance have to ascertain not only that the best physical theories did not preclude this account of causality from the start, as they would, say, were energy not a well-defined quantity in them, but to ascertain also that the particular types of interactions demanded by such an account were representable in the theories—that the theory supports the existence of energy and predicts that it will be exchanged in certain situations does not *eo ipso* guarantee that energy will be exchanged when and how a particular such account of causality requires. Best of all would be an argument that showed how such an account of causality could be "read off" directly from the mathematical or conceptual structure of our best theories, in the same way as classical particle mechanics is thought to require no interpretation.

3 Causal Relations and Energetic Processes

To make a start on sorting out the different ways energy can plausibly be thought to bear on causal relations, consider that exemplar of causation, the naive picture in classical physics of the motion of impenetrable, perfectly elastic bodies and their impinging on each other—Hume's game of billiards, say. The usual story says that when such a body in motion, A, strikes such a body at rest, B, under "normal conditions", the first body imparts motion to the second by transferring to it some of its kinetic energy: A's striking B causes B to start into motion. The first step in the analysis of the possible relations of energy to causality will be to tease out of this brief telling of the old chestnut the propositions that warrant the causal claim.

Assume that A and B are completely isolated from other physical systems—perhaps they are floating in the near-perfect void of intergalactic space—which is to say, they are not interacting with any other dynamical systems in such a way as sensibly to affect their dynamical evolution. A few quadrillion photons may be hitting them, a few quintillion neutrinos passing through them, but it is an excellent approximation to treat them as free bodies; in particular, they themselves are slight enough so that any gravitational force they may exert on each other or on themselves may be safely ignored. Assume we are observing the system in a laboratory co-moving with B, so it appears to be at rest (recall that this example is taking place in Newtonian spacetime, so this assumption is viable). By hypothesis, before the collision A propagates with a uniform velocity so that its center of mass traces out a straight line that, continued indefinitely, would pass through B's center of mass. As A's center of mass successively moves along this line before the collision, A's kinetic energy approximately equals its total energy, which remains unchanged—it is conserved. This is one of the indications that A indeed is not interacting with any other dynamical system, though it seems that this is only a necessary and not a sufficient condition. We would likely want to say, for example, depending on the nature of the investigation, our purposes in studying the system, the sort of machinery we prefer to employ to model such systems, *etc.*, that a ball swung in a circle at a constant rate on the end of a string interacts with the string and with whatever holds the string. In this case, we may predicate the claim that the ball interacts with the string on the fact that, as formulated using Newton's second law of motion, in order to solve the ball's equation of motion, one must know the force the string exerts on the ball.[12] Still, because this force acts everywhere perpendicularly to the ball's acceleration—the string performs no work on the ball, and vice-versa—the ball exchanges no energy with the string.[13]

Now, B's center of mass does not start into motion immediately at the instant of contact, just as the state of motion of A's center of mass does not alter immediately at the instant of contact—this would happen only if the two bodies were perfectly rigid, but no bodies in fact are.[14] The shock waves the impact generates in the bodies take some small but finite time to propagate and reverberate through the respective bodies, ultimately setting the center of mass of each into roughly uniform motion apart. More precisely (though not much more), when the two bodies come into contact the bit of B's surface that gets hit is deformed ever so slightly inward under the force exerted on it through the rapid deceleration of the bit of A's surface hitting it. This deformed part of B exerts a force as it bends inward on bits of B further inside, as those bits inside resist altering their relative position in virtue of their stable cohesion, and so on, all throughout B (and similarly for A). Mechanical work is defined as a force acting through a given distance in the same direction as the force points, so in the first instant the bits of A in contact with B do work on B; in the second instant the bits of B deformed inward do work on bits of B further in towards its center, and so on. By hypothesis again, the bodies are rigid enough so that it is an excellent approximation: to set to zero the time between the instant of contact and the instants at which the states of motion of their respective centers of mass alter; to set to zero the amount of kinetic energy converted by the impact into gross internal vibratory motion of each body as it flies away; and to set to zero the amount of kinetic energy that gets transformed into thermal energy in both A and B through the random thermal fluctuations excited in their constituent particles during the time the waves of deformation pass through them immediately after contact. The net result is that A's direction of motion changes and its speed of propagation diminishes, and that B starts into motion with precisely the direction and speed to compensate for A's lost energy and momentum: the sum of A's and B's kinetic energies after impact equals A's kinetic energy before impact, and the vector sum of A's and B's linear momenta after impact equals A's linear momentum before impact, and similarly for angular momentum defined relative to any arbitrarily fixed point.

Were one inclined, say, to a point of view similar to Hume's, one could not draw any *causal* conclusion from this description of the process that did not ultimately depend on the way past experience had habituated one's reasoning faculties; if one followed Al-Ghazzali, on the other hand,

[12] This is not true, for instance, in the Hamiltonian formulation of the ball's equation of motion.

[13] We are ignoring such niceties as the change in the string's thermal and internal electrostatic energy induced by its being stretched.

[14] Indeed, in relativity theory one cannot even formulate a notion of absolute rigidity. One has only so-called Born rigidity (see, *e.g.*, Pauli 1921) as the best approximation to the idea in classical physics.

one would conclude that the will of god served as the only guide keeping the behavior of these phenomena, as manifested at each instant, in conformity with the regularity we ascribe to causation, though in fact we would think that no such relation obtained between any different events or bodies *per se*. In either case, one could say nothing about any possible relation among dynamical systems, intrinsic to them alone, not depending in any way on the way god's will intercedes at every moment to impose structure on the world, and not depending on the history and constitution of one's own mental faculties, in virtue of which one (state of a) dynamical system might justifiably be said to be the cause of (a state of) another.

For one who champions the idea that causality is a feature of the physical world, however, not reliant on the continual intercession of a god and independent of the history and constitution of our mental faculties, the story offers an obvious option: A's striking B caused B to start into motion in virtue of the transfer of some of A's energy to B via the work performed by mutual pressure during the time they were in contact. The crucial supposition here is that the energy manifested by B as it starts into motion "came from" A, in some sense or other: A imparts some of *its own* energy to B during the process. Were A to "strike" B and B subsequently to start into motion, though one were able somehow to determine that the kinetic energy B suddenly had acquired came not from A but had been transmitted in some immediate, occult fashion from the distant body Q (and perhaps that the kinetic energy that A had lost as it slowed or stopped upon contact with B had instantaneously been transferred to the distant body Z), or even that B's newfound kinetic energy had simply sprung spontaneously into being at that moment *ex nihilo*, and likewise that A's lost energy simply vanished *ad nihilum*, one would naturally conclude that A's coming into contact with B had not been the cause of B's starting into motion, but rather the transfer of energy from Q or the spontaneous creation of energy had been.

No sane person, I wager, could bring himself to swallow the Gargantuan global conspiracy required for these sorts of events always to be occurring in accord with the principle of the conservation of energy; for, with no physical theory to account for exactly which other bodies Q and Z might be involved in the transferrals of energy and why, or to pick out the conditions under which energy might vanish and appear spontaneously, a global conspiracy must be the only explanation. The total conservation of energy during the process, predicted theoretically and verified experimentally, serves, it is claimed, to justify the idea that B's newfound energy once had been A's, and it is this tangible link between the two that is to provide the ground for the causal claims one wants to make about this process: that A's propagation along that path, a manifestation of its kinetic energy, caused it to come into contact with B; that the impact of A and B caused both B to start into motion *and* A's state of motion to change;[15] that A's state of motion at earlier times was the mediate cause of

[15]Note the peculiarity of this instant in the proceedings, that one is forced to make two apparently distinct causal claims about it rather than only one. This reflects the fact that the impact of A and B mediates an *inter*action of the two systems, and not merely the unidirectional action of one "active" system on another "passive" one, so to speak. Philosophers have often ignored this interactive aspect of physical phenomena, which has led to much confusing and confused ink spilled on the question of the 'directionality' of causality. See Stein (npub) for a more thorough discussion of closely related topics.

B's state of motion at later times. Without the idea that B's newfound energy once had been A's, there seems no way to bind the two into a relationship intimate enough to move beyond the simple conditional that mathematical physics provides—'If A strikes B thus, A and B will move so'—and into the *recherché* realm of the causal.

What roles, then, schematically stated, must energy play in order to support causal judgements? Its primary job is to provide warrant for asserting judgments of the form "C causes E" when the solution to the equations of motion has already affirmed the truth of the proposition 'if C then E'. The C's and E's, as shown by the example of this section, will in some cases be different states of the same system (and so will necessarily obtain at different times), as when the state of A above at the time it contacted B is said to have been mediately caused by its state at earlier times. In other cases, they will be states of different systems at the same time, as when the state of B at the time it started into motion is said to have been caused by the state of A at that time. Finally, in still other cases they will be states of different systems at different times, as when the state of A at some earlier time is said to have been the mediate cause of the state of B at some later time. In the first type of case, it is the continuous propagation of energy, here in the form of the continuous motion of a ponderable body, that is to warrant the causal claim; in the second, it is the proximate transfer of energy, here in the form of the work performed by contact pressure between two ponderable bodies, that is to do so; and in the third, it is a combination of the two. I therefore turn now to analyze these two types of energetic processes (assuming that the third can be understood as a simple combination of the first two) to determine the properties energy must have to realize these processes.

4 The Propagation and Transfer of Energy

Consider first propagation. There is no such thing as pure energy propagating, of and by itself, as its own entity—as its own dynamical system. To use a scholastic idiom for a moment, energy is not a substance to support attributes. Even photons, which may appear to be so, are not: they have also momentum, angular momentum and spin. One could with the same justice say that the photon was pure momentum, with the attributes of energy, angular momentum and spin, as say that it was pure energy with the attributes of momentum, *etc*, and, in any of these cases, it would not be just at all. The photon itself is the system, and the rest are its attributes.[16] The famous relativistic equation of mass and energy might lead one to think otherwise, but it is beside the point, for mass itself is only one more possible attribute of a system and does not by itself constitute a substance capable of supporting attributes. 'This material thing here, with a certain mass, has this momentum' is a meaningful scientific proposition, but not 'this quantity of mass here has this momentum' (excluding the colloquial use of 'mass' to refer to a material system). Talk about the propagation of energy must always be understood to be shorthand for talk about the propagation of a particular dynamical

[16] One should take this "substance/attribute" talk with a pinch of salt. Even in classical physics, for instance, it is likely more proper to represent quantities such as momentum as relations rather than as attributes—a relation, say, between a dynamical system and an orthonormal tetrad on a region of spacetime. The scholastic jargon I hope only drives home more forcefully the point I am trying to make.

system, to which the energy is attributed.

In order to conceive of energy as propagating, therefore, one must assume a physical system that propagates, to which the energy is attributed. In order for the propagation of this energy, as carried by the system, to support the desired kinds of causal claim, we must know what it means, in classical physics, to identify a physical system, at one point of space at one moment of time, as being the same in a substantial sense as a physical system at a different point of space at a different moment of time, and, in relativistic physics, to do the same for systems at different points of spacetime. (From here on, for brevity's sake I will speak generally of points of 'spacetime', in the contexts of both classical and relativistic physics, distinguishing the two cases only when a crucial point rests on the difference in spatiotemporal structure between the two.) Besides knowing that the system bearing the energy is the same over time, in some substantive and relevant sense, moreover, one must also know that the system is of an appropriate sort to exchange energy with the affected system, with respect to both the amount and the rate of exchange required, which means that one must have a classification of physical systems based on some substantive and relevant differentiæ on the basis of which one can found such judgments. In somewhat more technical terms, one must know, for a given kind of physical system, the types of coupling with other systems (physical interactions) it can manifest, with what other sorts of systems those couplings can occur, and what the strength of such couplings may be in all the possible cases (how much energy will be exchanged, at what rate, in which direction, *etc.*). In our example of the balls, for instance, we are confident that the quintillions of neutrinos passing through B at the moment it leapt into motion had no role in this effect. We know this not because the neutrinos do not carry enough total energy to do this—in fact, they do—but rather because the neutrinos do not interact with, do not couple with, ponderable bodies in a way that has as part of its character that kind of exchange of energy. We know this, furthermore, because of the knowledge we have of the type of system a neutrino is, and, in particular, the knowledge we have of the admissible couplings, and the strengths of those couplings, neutrinos manifest with other types of systems.

In sum, in order to use energetic quantities and processes as the foundation of an analysis of causal relations, we require at a minimum the following capacities:

1. to identify a physical system as belonging to particular category of physical systems
2. to reidentify an individual system over time, in the course of its dynamic evolution
3. to quantify and measure to the required degrees of accuracy and exactitude the values of the physical quantities borne by the system we hold responsible for the causal efficacy in any given case

With this knowledge in hand, the idea is, we can rely on the quantitative agreement in the total amount of energy as it distributes itself among different systems in the course of particular interactions, as guaranteed by the principle of the conservation of energy, to give sense to claims of the form 'the transferral of energy *from A to B* at the moment of their contact caused B to start into motion'.

In both classical and relativistic physics, the identity over a spatiotemporal interval of a dynamical system is constituted by (at least) the continuous occupation of the points of the interval by an entity the same in all (or enough) relevant respects. This is not the vacuous truism it may seem. To see why, fix a given physical entity. Its state at a "particular instant of time" is represented by a bounded region of a spacelike hypersurface ("the region of space the body occupies at that instant"), with a particular attribution of values, in that bounded region, for the physical quantities it bears.[17] A set of equations of motion determines the appearance and behavior of the entity at later times. With very few exceptions, all known sets of such fundamental equations of classical and relativistic physics have the following character. Starting from the given values of the entitity's properties in the initial region of spacetime, the equations will have a unique solution for at least some finite time into the future. Based on this solution, one can construct a continuously varying family of continuous, mutually disjoint, timelike paths, such that

1. the paths collectively form a four-dimensional, solid, spatiotemporal tube, through every point in the interior of which passes exactly one path in the family
2. each path in the family has a point in the initial bounded region as its endpoint, and every point in the bounded region is the endpoint of exactly one of the paths in the family
3. if one parametrizes the curves by proper time,[18] and one fixes some small enough positive real number δ, then the collection of points consisting of the point on each path a time δ later than that of the initial region forms a bounded region of a spacelike hypersurface ("the region of space the body occupies at that later instant"), and the physical system occupying that region is of the same type as the system in the initial region, in the sense that one can characterize its state using the same physical quantities, and the same set of equations of motion determine its appearance and behavior at later times

This is just to say that, in their guise as differential equations, the system's equations of motion possess well-posed initial-value formulations.[19] It is this predictably homogeneous aspect of dynam-

[17] We are glossing over many questions of interest and difficulty, that are beyond the scope of this paper. For example, it is not clear to me that one can consistently and coherently frame a precise notion that captures what we seem to be gesturing at with the idea of "all the physical properties borne by a physical system". Consider, for example, a body of viscous fluid. What are "all its physical properties"? Those such as its temperature, gross fluid velocity, its hydrostatic pressure, its state of shear and internal stress, its coefficient of thermal conductivity, and all the rest one needs to model the theory using the Navier-Stokes equations? Must one also include the acceleration field of all its constituent molecules? The quantum state of each of the atoms in each of its molecules? The SU(3) representation of all the strong-force interactions occurring in the nucleus of each of those atoms? I do not think that these are questions that can be answered in the abstract. One must rather have the context of a particular sort of physical investigation, employing a specific set of theoretical and experimental tools, in order to make the fundamentally pragmatic decision about the way one will model the system, including the physical quantities one treats as attached to it during the time one models it. See Curiel (2011) for a discussion of these issues.

[18] This means that any observer whose worldline instantiates such a path will, over the course of any segment of the curve, record an interval of proper time as having passed numerically equal to the metrical length of the segment.

[19] See Geroch (1996) for a thorough discussion.

ical systems, a guaranteed consequence of the *form* of their equations of motion, that ultimately underwrites our identification of them over spatiotemporal intervals.

This last remark, on the fundamental role played by the form of the equations of motion, demands an explication, by way of a brief detour. Fix, for the purposes of this detour, a physical system and the system of partial-differential equations it obeys. An initial-value formulation of the system consists of an attribution of values to all the dynamically relevant quantities of the system—at least, "all" with respect to the theory modeling the system using the partial-differential equations at issue—in such a way that the partial-differential equations have a unique solution for some finite interval of time, and a solution, moreover, that is stable in the sense that small (in a technical sense) perturbations in the initial conditions yield small (in a technical sense) differences in the induced solutions. For example, one could with equal justice say that both the mean kinetic energy of the molecules of a Navier-Stokes fluid and its gross, fluid temperature are, in some sense, "dynamically relevant quantities" it possesses. If one is using a thermodynamical theory to model the fluid, however, say one comprising the classical Navier-Stokes equations, one will treat the temperature and not the mean kinetic energy of the molecules, whereas using a statistical theory to model the fluid, *à la* Maxwell-Boltzmann, one will treat the mean kinetic energy of the molecules and not the temperature.[20]

Now, the equations of dynamical evolution most often thought of as "fundamental" by the physicist are of a particular mathematical type[21]—they are quasi-linear and hyperbolic.[22] Hyperbolic equations have three particularly nice, inter-related properties for the representation of the dynamical evolution of physical systems not shared by other types of partial-differential equations:

1. their characteristic "wave-fronts" propagate at strictly bounded speeds (the maximal speed depending on the particulars of the equations at issue)

2. their solutions are not necessarily analytic functions

3. discontinuities in initial data propagate in a continuous fashion through the solution to an initial-value formulation using that initial data

[20] See the third paper in this dissertation, "On the Formal Consistency of Theory and Experiment in Physics", for an expanded discussion of this point.

[21] I am skeptical of the idea of "fundamental" theories in physics. Consider quantum field-theory. It can not solve in closed form the dynamical equations representing the evolution of arguably even the simplest micro-system, the isolated hydrogen atom. It rather relies on perturbative expansions, and thus requires the system to be not too far from equilibrium of one sort or another. Quantum field-theory in general, moreover, can not handle phenomena occurring in regions of spacetime in which the curvature is too large. The Standard Model breaks down in regimes far above the Planck scale. Not even quantum field-theory formulated on curved-spacetime backgrounds can deal rigorously with phenomena under such conditions. I know of no theory of quantum gravity mature enough for it even to attempt the claim that it can be thought of as fundamental. And when, as I sincerely hope, one of these theories does mature and gain primacy, it will have no more warrant for proclaiming itself the ultimate bedrock than any of its predecessors.

[22] See, *e.g.*, Sommerfeld (1964) for a discussion of this type of equation, as well as the parabolic and elliptic types, with regard to how their respective properties bear on the modeling of physical systems. I discuss in the third paper of this dissertation the privileged role hyperbolic equations appear to play.

The first implies, among other things, that one can guarantee that the system modeled by the equations propagates at speeds less than that of light, as demanded by relativity. The second means, roughly speaking, that arbitrarily distant systems can not "influence" its evolution.[23]

The third requires a more involved explanation. Consider a metal rod at a uniform temperature, 10^{o}C higher than that of the ambient environment, say, air. If one represents the temperature of the entire system, the rod plus its environment, by a scalar field, in the context of a grossly thermodynamical theory, then that scalar field will have a discontinuity on the two-dimensional, spatial surface determined by the boundary of the rod—it leaps (or falls, depending on which way one is going, so to speak) by 10 degrees at all points on that boundary. If one models the subsequent thermal evolution of the system using a non-hyperbolic partial-differential equation, say, Newton's law of cooling, which is parabolic, starting from these initial conditions, then at *any* finite time after the initial instant the solution representing the joint state of the system, rod *cum* environment, will be an analytic function. The discontinuity in the initial data has been smoothed out and the values of the temperature and its derivatives at any point of the system equally determine the value of the temperature and its derivatives at all other points of the system, instantaneously. If one uses a hyperbolization of Newton's equation,[24] formed, for example, by adding to its lefthand side terms purporting to represent "relaxation effects", perhaps in the person of higher-order moments of the distribution function, then, in the absence of other interferences, at any finite time after the initial moment, the solution will exhibit a discontinuity in the value of the temperature, near the spatial boundary of the rod. The measure of the discontinuity, moreover, the scalar field on the boundary representing the jump in value of the temperature, will, in general, itself be a continuous and continuously varying field. I'll refer to this sort of case as the *propagation of a discontuinity in value.*

This system bears another, more subtle, and perhaps more important, possible discontinuity, closely related to the one just discussed: a discontinuity in the matter and the form, as it were, of the partial-differential equations used to model the two parts of the system, rod and air, in the way most appropriate for the requirements of the investigation at hand. Consider first a discontinuity in the matter only. In this case, equations the same in form—both being instances of Newton's law of cooling—model the air and the rod respectively. The two equations differ in the values of the constant, kinematic quantities characterizing the two systems, in this case the thermal conductivity of each, and the value of this quantity once again jumps discontinuously at the boundary of the rod. One may find it more appropriate, in the event, to use the Navier-Stokes equations to model the

[23]It is worth remarking that it follows from this fact—that only hyperbolic equations can have non-analytic solutions—that, contrary to what is often opined by both philosophers and physicists (see, *e.g.*, Russell 1927—though it is at least excusable in his case, in so far as the classification of partial-differential equations in this way, along with the clarification of their properties, was only then being settled at around the same time as he was writing, by, notably, Hadamard 1923), the condition that physical systems obey a "principle of causal locality" is *not* expressed by the fact that equations of motion are partial-differential equations *simpliciter*. It is rather expressed in the fact that equations of motion tend to be *hyperbolic* partial-differential equations. If they are not hyperbolic—for instance, if they are parabolic, as are the classical Navier-Stokes equations—then all bets are off about the "locality" of interactions.

[24]See Geroch (1996) for a discussion of hyperbolizations.

thermal evolution of the air, while still using Newton's equation to model the thermal evolution of the rod. In this case, not only will the system possess a kinematical discontinuity—the "matter" of the partial-differential equations changes discontinuously—but it will have a dynamical discontinuity as well—the partial-differential equations change in form across the boundary. This discontinuity also propagates in a continuous fashion.[25] Consider, in constrast, a discontinuity in the value of the temperature of the rod itself. At a given moment, say, the left half of the rod is 10^oC hotter than the right, and this difference manifests itself as a discontinuity in the value of the temperature in the interior of the rod as one passes through a lateral surface moving from the one side to the other. In this case, the system manifests neither a kinematical nor a dynamical discontinuity—it is continuous in both matter and form, suffering a discontinuity only in the value of one of its attributes.

We are finally in a position to offer a tentative characterization of the temporally continuous identification of a physical system as being the same, in those respects normally germane to physical identification as determined by the requirements of the investigation at hand. The interior of a four-dimensional spatiotemporal tube represents the course of dynamical evolution of a single, continuously identical physical system if

1. the partial-differential equations modeling the dynamical evolution of the physical fields one's theory ascribes to systems of that type suffer no dynamical or kinematical discontinuity within the tube

2. the values of these fields on any spacelike slice through the tube constitute a well set initial-value formulation of these equations

3. the tube is maximal in the sense that no tube containing it satisfies these conditions

I must stress that this characterization pretends to offer only necessary, not sufficient, conditions for the temporally continuous identity of physical systems. One must account for many pragmatic factors as well, if one wants to classify the system as being of a single, well recognized, continuous type over the course of its evolution, such as 'photon' or 'Hydrogen atom' or 'pendulum', for many types of systems that we think of as different in many ways obey, in some theories treating them, partial-differential equations identical in form. The three cases just listed, for instance, can all be modeled as (superpositions of) simple harmonic oscillators.

Note that what counts as "those respects of the system germane to identification" will vary from type of system to type of system, and even for the same system as it appears in different sorts of investigations, in so far as different theories will be used to model the dynamical evolution of the system in different investigations. Were one, for example, able to track a photon traversing cosmological distances through an expanding cosmos such as ours, the fact that its energy continually decreases[26] (the "red-shift effect") would in many circumstances not stop one from asserting that it

[25] One can make precise these ideas about discontinuities in the form and the matter of partial-differential equations by treating them as quasi-linear operators over appropriate function spaces.

[26] It is delicate to state this a precise proposition in the context of general relativity, but it can be done. See Wald (1984, §5.3, pp. 101–4), for example.

was in an important sense the 'same' photon that got emitted from a certain star. Otherwise one would appear to rule out the possibility of investigating dynamical systems at cosmological distances from us: if one cannot assert that this photon is in some important respect the 'same' as the one emitted from that star, one will normally have no ground for using any information gleaned from the photon to infer any information about the star. In other cases, changes in a system concomitant with changes in its energy may very well push one to conclude that the resulting system is not the same as the original system. If one pumps enough energy into a Hydrogen atom, for instance, eventually the electron will escape the central proton and fly off freely. To an organic chemist, the widely separated, relatively independent proton and electron may no longer constitute the same system as the original hydrogen atom, whereas a high-energy particle physicist may consider them to be precisely the same system, only in a very different state than before. Finally, in some instances there may be no way to reidentify a system indefinitely over spatiotemporal intervals: I know of no way to pick out 'part' of an excited atom and rightfully assert that it is identically continuous with the photon the atom absorbed a moment earlier.

I take it as a fundamental assumption of physical science that "those respects of the system germane to identification" can be defined in any given case, even though there is no universal formula specifying a procedure for doing so in all cases. There is, for instance, no variable in its equations of motion the value of which represents the fact that the system is, say, a hydrogen atom—only variables for position, momentum, *etc.* That what is being modelled is a hydrogen atom is encoded in the formal relations among the variables representing its dynamical quantities and in the values of the intrinsic, kinematic parameters one must fix (mass, spin, *etc.*) to represent it, *i.e.* precisely in the form of its equations of motion and in the canonical geometry of its space of states and the set of vector fields on its space of states representing the system's kinematically allowed dynamical evolutions.[27] Let us call the set consisting of the system's equations of motion, its space of states, and the set of vector fields on its space of states representing its kinematically allowed dynamical evolutions the *dynamical representation* of the system. Then the similarity in form of the dynamical representations of entities at neighboring spacetime regions is in almost all cases a necessary condition for identifying the entities as being spatiotemporally proximate manifestations of one and the same dynamical system. Let this suffice for a discussion of the first condition required for reidentifying a bit of energy over spacetime intervals, that of the reidentifiability of dynamical systems over spacetime intervals.[28]

The last of the three examples cited above, that of the atom that had absorbed a photon a moment earlier, suggests what is required for the second condition, that or those guaranteeing that the energy of the evolving, continually identical system remains the 'same' in the sense required to support causal claims. It is not the case that a hunk of energy can be identified once and for all, no

[27]This point is related to the remark of Stein (npub, §VI, p. 15), to the effect that the fundamental forces of physical theory are most aptly analogized not with the Aristotelian efficient cause, but rather with the Aristotelian formal cause.

[28]I cannot stress enough that this discussion as it stands is far from adequate. Spatial and temporal constraints do not allow a proper treatment of the issue. I hope to publish one in the near future.

matter what happens to a system that happens to 'contain' it at any given moment, so to speak. A hunk of energy can be identified over time as being the same in the relevant sense *only* so long as its associated system evolves in isolation, with no external interactions; moreover, one cannot naturally 'divide up' the entire energy content of an isolated system into separate parts in order to keep track of such parts over time. One can keep track of and reidentify only the entire quantity of energy associated with an isolated system over time. What is needed then is a criterion for determining when a system is 'isolated', which is to say, not interacting with its environment.

The analysis of the identity of dynamical systems just offered, as depending on the form of a system's dynamical representation, suggests a definition of 'in isolation'. A system will be said to be 'isolated at an instant' if its actual equations of motion at the instant imply conservation of all classically conserved quantities. The system will be said to be 'isolated during a spatiotemporal interval' if it is isolated at every instant of that interval. Stipulating that the system be isolated, though, does not by itself suffice for concluding that the energy associated with the system at each instant it is isolated is in some significant sense the 'same'. The forms of the equations of motion of isolated systems both in classical physics and in special relativity certainly imply that isolated systems will have a definite quantity of energy at every instant, as represented in a fixed global coordinate system, but they do not state that the energy of a given system is the 'same' in any way at any two instants other than perhaps being quantitatively the same. Again, there is no variable in the equations of motion that labels a particular bit of energy and whose time-derivative tracks its evolution. Identifying the energy of an isolated system (in both classical physics and special relativity) as the 'same' during the period it is isolated is justified by the fact that the energy of an isolated system is conserved: the quantity of the system's energy at each instant is the same. This brute fact is supposed to justify the thought that energy can be neither created nor destroyed, and so, *a fortiori*, is the 'same' at each instant in some physically significant way. One must not take 'identifying a hunk of energy associated with a system' in too strong a sense, even for periods when that system is isolated. An isolated, excited atom, for example, may at some point emit energy in the guise of a photon, but there is no sense that can be attached to the question, *which* particular bit of the atom's initial energy was emitted in the form of the photon. This is the force of a remark by Maxwell (1877, ch. VI, §109, p. 90), which might otherwise be taken to controvert my discussion here: "We cannot identify a particular portion of energy, or trace it through its transformations. It has no individual existence, such as that which we attribute to particular portions of matter." What one *can* posit is the following proposition, required by accounts of causality that wish to invoke energetic processes to support their causal claims: that a certain unchanging quantity of energy is identifiably 'attached' in a significant way to a particular dynamical system evolving in isolation, while it so evolves, so long as the equations of motion of the system imply the principle of the conservation of energy, as do those of isolated systems in classical physics and special relativity.

With so much behind us, we need not take long discussing the second of the processes required for supporting causal claims, energy transfer from one system to another—that, when two (or more) systems interact, the gains and losses of energy of each system during the interaction can in a significant way be matched up with each other, *e.g.*, that the energy gained by A was transferred from B, or the energy lost by A was transferred in part to B, in part to C and in part to D, *etc.* The same sort of analysis as worked for propagation will apply here as well. Consider the interaction of two dynamical systems. Barring Newtonian gravity, the fundamental interactions of systems in classical and relativistic theories share this feature: two dynamical systems, whether both are classical or both are relativistic, are represented as interacting with each other only when, roughly speaking, at least some bit of one is spatiotemporally contiguous with some bit of the other. Except in examples such as that of the ball swung on a string at a constant rate in a circle, in which the magnitude of none of the velocities of the systems changes during the interaction, the interactions of systems in classical and relativistic theories also share this feature: during the interaction, the energetic quantities of both systems will jointly alter in a regular, predictable way. Both of these features, again, are consequences of the form of the equations of motion the systems obey during their interaction. Using the equations of motion to represent an interaction during which energy is conserved, *e.g.*, one can predict that a decrement of energy in one system will be exactly and simultaneously counter-balanced by an increment of energy in a system in some way spatiotemporally continuous with the first. This regularity and predictability of changes in energy in adjoining systems, which itself is guaranteed by the form of the equations of motion, *partially* warrants the claim one needs to support causal relations in the desired way: that the systems have *exchanged* energy—that the one has lost and the other has gained, in virtue of the fact that they were interacting, the 'same' energy. This regularity and predictability by themselves justify only the statement that energy changes in interacting systems are correlated with each other. To warrant the further claim, that energy is exchanged, a criterion is still needed for determining when energy actually 'passes' from one system to the other—when one system has *acted on* the other. Otherwise one is still in the realm of the initial value formulation of ordinary differential equations and the simple conditional propositions it entails, but not in the purportedly richer realm of causality.

Now as remarked above, one cannot tag hunks of energy as one can hunks of cheese, and so one cannot identify the energy that this system lost with the energy that that one gained in the same way one could if one were talking about cheese. The way it is actually done in scientific practice relies on the fact that, in interactions represented in classical physics and in special relativity, it is a consequence of the form of the fundamental equations of motion that energy is conserved. If one considers the physical concatenation of the interacting systems itself to be a single system, and this combined system is isolated during the interval in which the original systems interact, then at any given instant the total amount of energy of the combined system is the same as it was at the beginning of the interaction. As a corollary, the rate at which one of the original systems gains energy during the interaction must be the additive inverse of the corresponding rate of the others, where a negative rate of gain represents loss. This brute fact provides the warrant for asserting that, in an interaction of the type considered here, the energy that one system loses is the 'same' as the energy the other system gains, at least in the way required to support the desired sorts of accounts of causality. If energy were not conserved in interactions, then the natural conclusion would be that some of the total energy of the two systems had been either created or destroyed during the interaction. If energy were the sort of thing capable of being created or destroyed, then in any given case in which the equations of motion asserted, say, that one system lost energy and another one gained it, there would be no more reason to infer that there had been *transfer* of energy from the one to the other than to infer that some of the 'original' energy of one simply had vanished and some 'entirely new' energy had simply appeared in the other one, with no other relation between the two at all—certainly no relation rich enough to support a causal claim that is supposed to assert more than would a bare conditional statement of the form: "If energy gained here, then energy lost there." The fact that it is always the same *amount* of energy that is gained and lost by interacting systems is supposed to preclude the idea that energy can be created or destroyed, and so warrant the inference that energy is actually *transferred* between interacting systems, as required.

In sum, a necessary condition for characterizing both the propagation and the exchange of energy so as to be of use to transfer accounts of causality is that dynamical systems satisfy the principle of the conservation of energy[29]—the form of law Bob Geroch in conversation referred felicitously to as "the mathematical representation of thinghood," in so far as it encodes traditionally essential features of substance such as its identifiability over time and its permanence, which is to say the impossibility of its creation *ex nihilo* and of its destruction *ad nihilum*.

There are actually two separate formulations of the principle of the conservation of energy, the differential and the integral. The one keeps track, at individual spacetime points, of the continuous *flow* of energy into and out of the immediate neighborhood, while the other compares gross quantities of the stuff in different bounded spacetime regions in timelike relation to each other. The differential conservation law guarantees that there are no sources (or sinks) of energy in the sense of the existence of a point into (out of) which more energy flows than flows out of (into) it. The integral conservation law guarantees that, if a dynamical system gains or loses energy, then that energy loss or gain 'registers' as a non-zero total energy flux through the bounding surface of some spatiotemporal volume completely containing the system, in just the proper amount to balance the amount of energy the system gained or lost. Another way to think of the integral formulation is that it precludes the existence of finite or gross, not necessarily localized energy sinks and sources.

The integral form of the principle is the crucial one for sustaining the idea that energy *propagates* and is *transferred*, and not merely that some appears here and some disappears there. Given a dynamical system possibly isolated for an interval, first draw a four-dimensional spacetime tube closely around the system throughout the interval. Applying the integral form of the law over this tube allows one to keep track of the system's energy for the whole interval, which includes determining whether the system gained or lost any energy during the interval and, if so, the specifics of the gain or loss. In so far as one has succeeded in reducing causal propositions in the first place to ones about energy propagation and transfer, this device further allows one to affirm propositions

[29]I emphasize that the satisfaction of such conservation principles is *only* a necessary condition for being able to define propagation and exchange of conserved quantities as needed for transfer accounts of causality. Howard Stein has argued convincingly in a private communication to me that the propagation of energy and of other classically conserved quantities *prima facie* do not always track the propagation of causally relevant information in, *inter alia*, classical electromagnetic theory. The propagation of optical information in a diffracted optical field, for instance, does not necessarily follow the flowlines of the flux of any classically conserved quantity in classical electromagnetism.

such as "The causal process evolved via *this* continuous chain of events, and not via any of those," rather than merely saying "The sum of events in this spacetime region here determined, or was causally responsible in some way for, the event or sum of events in that region there."

In classical physics and in special relativity, these two formulations of the conservation principle are essentially equivalent to each other—both hold in general and each implies the other—so all seems in place for the possibility that energetic processes as represented by these theories can be used to ground the desired causal claims.

5 Energy in General Relativity

General relativity does not naturally support any sort of mathematical structure with which to construct relations similar enough to classical conservation principles to wear the name gracefully, at least so far as the sorts of accounts of causality I am considering are concerned. The precise statement is that the only 'conservation law' one can formulate in a generic general relativistic spacetime is a differential *covariant* conservation law. No two physicists, not to mention philosophers, seem to agree on what exactly the import of being 'covariant' is for an equation in general relativity.[30] For my purposes, it suffices to say (which I think is not contentious) that a necessary part of what makes the differential covariant conservation law in general relativity covariant is the fact that the 'differential' in this law comes from the covariant derivative operator naturally associated with the ambient spacetime metric. Consequently, in a generic general relativistic spacetime there is no privileged, physically significant way to cast the differential covariant conservation law into the form of an ordinary partial differential equation or set of such equations, and so such a law cannot in general be transformed into an *integral* conservation law. There simply are no integral conservation laws in the generic general relativistic spacetime. If one accepts the argument I made in §4 above, that integral conservation laws are a *sine qua non* of defining propagation and transfer of energy (or of any classically conserved quantity), in so far as one wants to have these ideas support rich causal claims, it follows immediately that one cannot formulate transfer accounts of causality in a generic general relativistic spacetime.

That the covariant differential conservation law does not imply an integral conservation law follows from these two facts: first, that the fundamental "energetic" quantity in general relativity (as in special relativity) is not a scalar function on spacetime but is rather a two-index covariant tensor field, the stress-energy tensor T_{ab}, which can be thought of as a linear map from ordered pairs of vectors on spacetime to real numbers; second, that generic spacetimes in general relativity have no preferred class of "frames of reference", as Minkowski space in special relativity has, *viz.*, those defined by classes of worldlines of inertial observers all at rest with respect to each other. The lack of these two structures collude to hinder the formulation of an integral conservation law.

In special relativity, one also uses a tensor, not a scalar, to represent energetic quantities, and yet one still can formulate integral conservation laws perfectly well there. By applying the ambient

[30] See Norton (1993) for a thorough review of the topic.

stress-energy tensor in turn to each of the canonical timelike Killing fields (symmetries of the metric) on Minkowski spacetime, one constructs a canonical set of scalar fields naturally thought of as energy densities, and formulates a distinct integral conservation law for each scalar field in the set. In order to construct a scalar from a stress-energy tensor in general relativity, two vectors are needed (recall that the stress-energy tensor is a linear map from ordered pairs of vectors to real numbers—to scalars). The vector tangent to the worldline of an observer provides an obvious and natural candidate, and indeed it is a fundamental fact about general relativity that the ambient total energy density experienced by any given observer at a given spacetime point is precisely the real number one gets by applying the stress-energy tensor at that point to the ordered pair each component of which is the vector tangent to the observer's worldline at that point. So far, this is precisely the same procedure followed in special relativity to formulate the integral law of energy conservation. To formulate an integral law in general relativity, therefore, it would seem that all that has to be done is to pick an appropriate family of observers, *viz.* a family of timelike curves, so that by applying the stress-energy tensor to their respective tangent vectors at every point in the region one will get a scalar field representing the total energy-density of all matter in the region as experienced by those observers, which can be used to formulate the integral law. In special relativity, families of timelike curves representing inertial observers all at rest with respect to each other play this role, and because such worldlines in Minkowski space have a few extraordinarily nice technical properties—summed up in the proposition that they are simultaneously geodesics and the integral curves of a Killing field—it turns out that the scalar field one gets in this way does yield a perfectly good integral conservation law. One oddity of the situation is that one will get not one but an uncountable cardinality of different integral conservation laws, one for each 'preferred time-frame' defined by a family of co-moving inertial observers. There is nothing inconsistent about this—each family of observers will experience energy, *et al.*, as being conserved in their own time-frame, and will be able to predict that all other families of co-moving inertial observers will have the same experience, though no two families will agree on the values of the ambient energy density and flux. In fact, this is true in classical physics as well: one will get different integral expressions for the conservation principle for each different Galilean inertial coordinate system as well.

The crucial difference between special relativity and general relativity, between, that is, Minkowski spacetime and the generic general relativistic spacetime, is that, not only will there not be a family of timelike curves that are all simultaneously geodesics and the integral curves of a Killing field, but there will not even be a family of timelike curves that are simply the integral curves of a Killing field. In fact, this last structure by itself suffices for formulating an integral conservation law. To formulate an integral conservation law in a given region of a general relativistic spacetime that has a timelike Killing field, one would pick a family of timelike geodesics filling the region ('inertial observers'—these can always be found); then, to construct the appropriate scalar field, one would apply the stress-energy tensor at each point in the region to the pair of vectors one component of which was the tangent vector to the geodesic from the fixed family passing through the point and the other component of which was the Killing field vector at that point (because the stress-energy tensor is symmetric, it would not matter how one ordered the two vectors); finally one would use this scalar field to formulate the integral conservation law. The properties of the timelike Killing field ensure that one will be able to formulate an integral conservation law for the resulting scalar, in analogy to those of special relativity. The scalar field that results can be thought of in a certain sense as the total energy density at a point, with two important caveats. First, this 'energy density', though constructed relative to a particular family of timelike geodesics ('inertial observers'), will not be the energy density that any actual observer instantiating one of the geodesics would measure using any standard experiment for measuring energy density. Second, it does not include any contribution due to 'gravitational energy', since this is not localizable in general relativity.[31]

Now, the presence of intrinsic curvature in the spacetime manifold does not by itself imply that there cannot be timelike Killing fields: there are solutions to the Einstein field equation that represent curved spacetimes with timelike Killing fields, and in these spacetimes integral conservation laws can be defined. That the generic general relativistic spacetime possesses intrinsic curvature, though, does make it extremely difficult for it to have Killing fields, timelike or not. The reason behind this, intuitively speaking, is as follows. If an observer were to embody an integral curve of a timelike Killing field, she would record an extraordinary fact: the metrical structure of spacetime, in a sense that can be made precise, would appear to her not to change in the slightest as time passed. At every moment of her proper time, spacetime would appear essentially the same as at every other moment. For this reason, timelike Killing fields are said to represent 'time-translation symmetries'. This property undergirds the Killing field's capacity to yield integral conservation laws—they provide a physically significant temporal background, so to speak, against which one can track the gross quantity of energy in a given spatial volume as it 'evolves' with respect to the metrical structure of the spacetime, which thanks to the symmetry implied by the presence of the Killing field can be taken as 'constant over time' in a certain sense.

The spacetimes in which Killing fields occur, however, are highly special and unphysical. Special, because a generically curved spacetime will not manifest such extraordinary symmetry, as one ought to expect: think of a 'generically curved' surface—perhaps a sheet of rubber that is distended and stretched at random—and it will manifest any sort of symmetry whatsoever only under rare circumstances, much more manifesting a perfect, global symmetry such as is embodied in a Killing field. Unphysical, because such spacetimes are unstable against arbitrarily small inhomogeneities: the smallest speck of dust the tiniest bit out of place in only one spot in the entire spacetime precludes the existence of a Killing field. It is only in such unphysically dainty spacetimes, by dint of the daintiness itself, that one can define a quantity that behaves enough like energy even to be tempted to call it that.[32]

[31] If the Killing field is spacelike, then one may get conservation laws for quantities analogous to linear and angular momenta as they appear in special relativity.

[32] Another class of special spacetimes, the so-called *asymptotically flat* ones, also admit two precisely defined energy-like quantities. One is most naturally interpreted as the total energy contained in the spacetime 'at a single instant of time', *i.e.* in a single spacelike hypersurface (see Arnowitt, Deser, and Misner 1962); the other represents the total amount of energy 'radiated off to infinity' at any given time (see Bondi, van der Burg, and Metzner 1962 and Sachs 1962). These are both global quantities, akin to the total energy of a dynamical system in classical mechanics, but with one very strange feature: they have no local analogues—there is no scalar or vector field in such spacetimes that

Even if general relativity does not allow the formulation of such conservation laws, and so does not allow the defining of a quantity like energy as it appears in special relativity, one may still wonder whether energetic quantities useful to transfer accounts can be defined in other ways. They cannot. The fundamental structure of general relativity by itself does not provide the appropriate setting for any localizable energetic quantity to be rigorously defined in any way analagous to how such quantities are defined in either classical physics or in special relativity. They just are not fundamental components of the theory as they are in classical physics and special relativity. I will not enter here into the technical details of this result; I will only remark that the heart of the matter lies in the impossibility of defining in general relativity a mathematical object that represents a local energetic quantity specifically associated with the 'gravitational field'—one cannot ascribe a local energy density to gravity, or really any localized energetic quantity to it at all with any degree of rigor.[33]

On the face of it, this is an extremely puzzling result, for it is not difficult to convince oneself that one can extract energy from the gravitational field—after all, energy is continually transferred from the moon's orbit to the oceans through the work done in the rising and ebbing of the tides.[34] The principle of energy conservation, moreover, seems one of physics' most dearly held principles. Its consequences produce the predictions that have confirmed our most fundamental quantum theories to mind-boggling degrees of accuracy. Engineers employ it constantly in designing the contraptions that, by and large successfully, house, feed, transport and entertain us. So what gives? The answer is that general relativity tells us that, rigorously speaking, there is no such quantity, but that in certain sorts of approximations one can recover a quantity that is naturally identified as energy. When the background curvature of a spacetime region is 'small', one may treat the region as being for all practical purposes flat, with the consequence that there will be an 'approximate timelike Killing field', and one may proceed to define energy as one does in the presence of a true timelike Killing field. That this approximation holds good in the region of the solar system explains how the idea of energy can be so useful to us, and appear so fundamental, when in fact one of our two fundamental theories says it is not. This procedure is actually doubly approximative, in that there is no precise definition of an 'approximate Killing field'—in practice, physicists wing it on a case by case basis, and this is appropriate for their tasks. For we, though, who investigate the would-be *fundamental*

one integrates up to get this 'total energy' (see Curiel 1996). Consequently, while these quantities are of great interest for purposes of calculation, they do little good for the advocates of transfer accounts of causality.

[33]The precise statement is that one cannot define a 'stress-energy tensor for the gravitational field': the only two-index covariant, symmetric tensors that are concomitants of the Riemann curvature tensor are linear combinations of the Einstein tensor and the metric, and these are not viable candidates.

Similar results about the indefinability of gravitational energy hold in Newtonian gravity, though the situation is somewhat better there in that, in special situations, one can define an energy density for the gravitational field, which one can never do in general relativity. I felt it necessary to bring the heavy machinery of general relativity to bear against transfer accounts of causality, and not rest content with the example of Newtonian gravity primarily because general relativity is the fundamental physical theory, not Newtonian gravity, and I am interested in constraining accounts of causality that have some pretense of being fundamental.

[34]See Bondi (1962) for a more detailed argument that one can extract energy from the gravitational field in Newtonian theory, and Geroch (1973) for such an argument in general relativity.

features of the world as represented by general relativity, approximations, no matter how good and no matter how well justified in certain experimental calculations and practical endeavors, have no relevance.

Even though the idea of 'energy', classically so dependent on conservation laws for its definition, in one sense disappears in general relativity, it does not do so completely. I think it would be more accurate to say that the idea of 'energy' *alters* in the transition from classical physics to special relativity, and again in that from special to general relativity. In the first place, although this is not often explicitly recognized, in classical physics there are actually two separate conceptions of energy, each with its own distinct proper mathematical representation: energy as the capacity to do work (closely related to the idea of potential energy), properly represented by a 1-form on the space of states of a classical dynamical system, the 'work 1-form', *i.e.* a linear mapping from vectors tangent to the space of states ('rate of change of the state of the system') to real numbers; and energy as the generator of the time-evolution of a system (closely related to the idea of kinetic energy), properly represented as a scalar field on the space of states in conjunction with a mapping from scalar fields to a certain set of vector fields on the space of states, those representing the kinematically possible dynamical evolutions of the system. When these objects satisfy certain conditions, then one can formulate the usual conservation laws, which quantitatively relate the two conceptions of energy by equating the total amount of energy gained or lost by a system to the amount of work performed on or by it during an interaction.

In special relativity, there is fundamentally only one energetic quantity, the stress-energy tensor. The relativistic equation of mass and energy requires a mathematical structure that will keep track of the fact that energy flux has momentum and that momenta contribute to energy flux—which is all neatly encoded in the person of a two-index symmetrical tensor, *viz.* the stress-energy tensor. One can derive from it analogues to the objects representing the two conceptions of energy in classical physics by fixing an inertial coordinate system and decomposing the stress-energy tensor into its energetic, linear momental and angular momental components. Whereas in classical physics there were two fundamentally distinct conceptions of energy, united only by the conservation laws and this only contingently, special relativity teaches us that there is only one underlying quantity, stress-energy, with some, but not all, of the characteristics of energetic quantities in classical physics. Notably, integral conservation laws of a certain sort can still be formulated in special relativity, so the gross energetic quantities displayed in such laws can be related in physically significant ways to scalar energetic quantities, which are always well-defined. Finally, in the shift to general relativity one retains much of the structure of stress-energy in special relativity, except this key aspect only: there are in general no integral conservation laws, and correspondingly there are in general no well-defined scalar energetic quantities of physical significance. In many situations of practical and theoretical interest, however, one can formulate approximate integral conservation laws and the correlative scalar quantities with many of the properties such structures have in special relativity.

Although the idea of energy and the sorts of fundamental energetic quantities extant do shift dramatically as one progresses up the ladder of theory, they do not alter beyond recognition, and in fact there are fundamental continuities, as I have tried to emphasize. I think this is absolutely important to realize—points similar to it often get overlooked in contemporary philosophical discussions of 'paradigm shifts', and the like. I think it will be helpful in making this point more clear to take a very brief look at some historical material. To my great surprise, the hero among the early proponents of general relativity (of those I have read with some care—I make no claim to have perused a large fraction of the significant historical literature in this area), at least with regard to having made a beginning of formulating a coherent and I think largely proper conception of the role of the stress-energy tensor and conservation principles in the theory, is Eddington. He has of the time perhaps the most sophisticated treatment of the role the stress-energy tensor of matter plays in general relativity, of the way it is introduced in the theory, and of the proper view to have of classical conservation laws and why they are not fundamental to the theory.[35] I take the liberty of quoting him at length (Eddington 1923, §59, pp. 135–6):

> . . . [W]e have. . . spoken of [the differential covariant conservation law] as the law of conservation of energy and momentum, because, although it is not formally a law of conservation [since it has no integral formulation], it expresses exactly the phenomena which classical mechanics attributes to conservation. . . . ¶ As soon as the principle of conservation of energy was grasped, the physicist practically made it his definition of energy, so that energy was that something which obeyed the law of conservation. He followed the practice of the pure mathematician, defining energy by the properties he wished it to have, instead of describing how he measured it. This procedure has turned out to be rather unlucky in the light of the new developments. . . . We find that [the stress-energy tensor] is not in all cases formally conserved, but it obeys the law that its [covariant] divergence vanishes; and from our new point of view this is a simpler and more significant property than strict conservation.

I want to focus on the conclusion of the quotation, allowing the rest to speak (rather more eloquently than I could) for itself: Eddington claims that *general relativity itself* teaches us that the differential covariant conservation law is the principle it is proper to expect to govern the behavior of whatever energetic quantities there may be in the world, and not a classical conservation law that could be transformed into an integral conservation law. He does not make an explicit argument explaining why this equation is the one that best captures, in the mathematical language with which one represents the relativistic world, the physicists' practice of making energy and momentum measurements and finding that to an extraordinarily high degree of accuracy on the surface of the Earth certain classical conservation principles hold, but it is easy enough to sketch out what I believe he had in mind.

What general relativity, "our new point of view," demands to be taken into account is that for the purpose of quantitative calculation we represent all such measurements of energy and momentum in particular coordinate systems, none privileged over the others, and that all such measurements

[35] *Cf.* Eddington (1923, esp. §§53-4 and §59, pp. 116–22 and pp. 134–7). Russell (1927, ch. 9, pp. 84–95), one of the first philosophers to examine relativity theory with a high degree of mathematical sophistication, also took especial delight in Eddington's analysis. We are in no position today to condescend to Eddington, either—part of the original impetus behind this dissertation was a fruitless search for a philosophically illuminating contemporary examination of energy and conservation laws in general relativity.

can be approximated as occurring over an infinitesimal region of spacetime, on the supposition that they are not occurring in regions of extremely high curvature. Consequently a *covariant* law must be formulated that expresses the fact that, infinitesimally, these conservation principles are observed to hold when expressed in any coordinate system: all observers, no matter their state of motion will agree that these principles, properly formulated, do hold. The differential covariant conservation law precisely encodes all this information. This is why it deserves the honorific 'conservation law' even though it yields no integral equation in general: in a certain sense, the covariant law becomes a classical differential conservation law in the limit of the infinitesimally small. Thence the deep continuity between the general relativistic energetic quantities and classical conceptions of energy—they represent very nearly the same class of physical processes and operations, and are used to model the same experiments and to make predictions about them, predictions that in many cases are well-nigh indistinguishable among the various theories; moreover, the mathematical structure of the former can be shown to 'contain' the mathematical structure of the latter, in the sense that the structure of the latter falls out of that of the former under certain natural approximations, using certain natural manipulations. The vanishing of the covariant divergence of the stress-energy tensor is "simpler and more significant" in general relativity than a classical conservation law would be for the simple reason that it is a well-posed statement utilizing only structure intrinsic to the theory, *viz.* the stress-energy tensor and the affine connection of spacetime.

In the original derivations of the field equation that bears his name, Einstein repeatedly relied on the principle of energy conservation in arguments that motivated and even 'proved' many essential propositions.[36] Most of the reasons he gives for relying on this principle do not look so strong once one considers its status in the complete theory. In this respect, energy conservation is analogous to Mach's principle and to the principle of equivalence—a guiding intuitive principle that helped inspire the construction of the theory, but whose classical formulation does not seem quite to hold in the final theory itself. This fact by itself, however, does not invalidate his arguments: the deep practical, empirical and theoretical continuities among energetic quantities in the various theories I have emphasized show why Einstein's arguments are so good, why they are so successful, for in an important way we are still talking about the same underlying structural features of the physical world. Only now the concepts we use to investigate and understand them have evolved; they have not discontinuously metamorphosed. Correspondingly, the physical theory we use to represent and model these underlying structural features of the world has also changed. The sorts of calculations and derivations one can employ a stress-energy tensor in have changed, for example; the mathematical framework within which one specifies a stress-energy tensor has changed; judgments about the propriety of certain sorts of approximations—what to include in calculations, what to ignore—have changed. And so on. This new physical theory, though, both on its own and in its intricate connections to past theories and their better understood concomitant concepts, provides the key to understanding the new concepts it has introduced.[37]

[36] See for instance Einstein (1916, *passim*) and Einstein (1984, *passim*).

[37] The too short discussion of this paragraph was inspired in large part by the discussions of similar matters in and Stein (1989) Stein (npub).

General relativity demands revision of the classical conceptions of 'energy' and 'conservation'. I believe that we have not yet fully come to grips with the revisions, and perhaps abandonments, it urges on us.

6 Causality after General Relativity

So where does all this leave us? It seems clear to me that it leaves us with no way to represent transfer accounts of causality within the fundamental structure of general relativity. Almost every aspect of general relativity, in fact, militates against this conception of causality. One can predict with great certainty the regularity of certain relations among energetic sorts of quantities in general relativity, but this by itself will not suffice to support the types of causal claims advocates of transfer accounts of causality want to make. In the absence of integral conservation laws of the proper sort, there is no reason to take such regularity and its sure prediction as expressing anything more than the bare mathematical assertions that they are—"If energetic quantities of a certain sort, in a certain amount, are here, in this situation, then energetic quantities of a certain sort will be there, in a certain amount."

A reasonable first reaction to my arguments might be to give up the old characterization of propagation and look for a new one better suited to general relativity's *mise en scène*. Since general relativity does not allow the rigorous definition of a localizable physical quantity that has the essential features of energy as it appears in classical mechanics, however, nor of any other classically conserved quantities, it is not clear what one would have propagate even could one devise a new definition of it.[38] As a last ditch attempt to salvage the notion of propagation, one might be tempted simply to take particles themselves as what, by propagating, support a transfer account of causality—one cannot rigorously ascribe energy, mass or momentum to a particle, but its mere continuous, self-identical existence along its path through spacetime surely ought to count as a perfectly good case of propagation, and surely such propagation can underwrite the sorts of causal claims people want to make. This looks to be perhaps a promising avenue until one realizes that it will never work. Strictly speaking, one cannot formulate the Einstein field equations in a mathematically sensible way for point-particle sources, as one can, say, for the Maxwell equations.[39] Even were this technical hurdle surmounted, a more serious problem confronts this proposal: point particles are only

[38] One could perhaps try to use entropy and entropy-flux to define propagation and causal continuity, since entropy can be rigorously defined and treated, at the macroscopic level at least, in general relativity (see, for instance, Tolman 1987, §§119–20). So far as I know, whether one can give a rigorous treatment of entropy at the atomic and sub-atomic level in general relativity is not known, and presumably must await advances in quantum field theory on curved spacetimes and quantum theories of gravity. See Wald (1999) for a recent survey of this problem.

[39] The technical reason for this is that point-particles would have to be represented by a mathematical object known as a distribution, which is essentially linear; the Einstein field equation, being non-linear, has no well-posed distributional formulation. Recently, Colombeau (1992) has developed a theory of so-called new generalized functions that can be viewed as a non-linear generalization of distributions. Although in this new framework one can make sense of a much wider class of metrics than one could in the past, one still cannot rigorously construct a metric representing a point-particle. See Vickers and Wilson (1998) for a recent survey of applications of Colombeau's theory to general relativity.

idealized entities, useful in certain sorts of approximations; nothing in nature answers exactly to the idea. To the best of our knowledge, there are only extended bodies and fields, perhaps only fields. Consequently point-particles cannot be utilized to ground an account of causality with pretensions of being fundamental.

Finally, neither extended bodies nor fields will yield by themselves any way to define propagation in general relativity: there is no way to single out any particular curve in a spacetime region occupied by a spatially extended object or field in such a way as to give one reason to claim that anything of significance propagates along that curve. Quantities such as energy and momentum serve this function in classical mechanics, but one cannot call on them here. One might try to use the 'propagation' of an entire extended body to try to underwrite the desired causal claims, but my analysis in §4 above of the continuing self-identity of dynamical systems shows that, in doing this, one no longer is relying only on the fundamental structure of general relativity. For any dynamical system, the equations of motion by themselves do not contain an 'identity variable'—dividing the world up into discrete, extended bodies is *not* a part of fundamental physics as captured by general relativity, but is rather tied up with our preferred way of doing physics, what Bob Geroch evocatively calls 'psychology'. It should also be emphasized that such a conceit is extremely artificial when one considers *fields* rather than extended bodies—it is difficult to know how to make sense of the idea of a discrete, bounded 'chunk' of field propagating *en bloc*.[40]

General relativity does not by itself suggest entities or quantities that one will want to characterize as 'propagating', no matter how one defines it. The very different structure of spacetime in the theory from that of spacetime in classical physics and in special relativity does not naturally suggest any sort of transfer account of causality, nor does it easily admit one. The only reason I can imagine for trying to force one to fit into the framework of general relativity is because one approached the theory in the first place already with a set of classical notions and questions to address, and did not rather ask general relativity what the important notions and questions ought to be in its new framework.

If one renounces transfer accounts of causality, as I see it there remain only two general sorts of accounts of causality that could be grounded in physical theory. One may postulate an account in which one or more discrete, localized entities that are spatiotemporally separated from each other 'cause' a distinct entity, the 'effect', itself spatiotemporally separated from all the 'causes'. Just as with any attempt to hold on to propagation, however, such an account would in no way arise from the structure of general relativity itself, but would rather have to be forcibly superimposed on its structure, under the guidance I suppose of purely metaphysical urges. Otherwise, there is the initial-value formulation of mathematical physics, my preference for the best one can do in representing causality in general relativity.[41] Whatever sort of account one will give looks to come perilously close merely to saying that one thing *follows* upon another.

[40] I actually should want to say that another lesson general relativity urges on us is that the distinction between fields and ponderable bodies is not a fundamental one, but that is a sermon for another time.

[41] Of course one also has the option of not giving an account of causality at all, and simply going about one's business with the physical theory—this may be my favorite of the options.

I have been concerned with accounts of causality that aspire to be fundamental, to reflect the actual, basic structure of the physical world as best we know it. I would not desire to preclude from the philosophical and scientific armory all notions of causality that depend on ideas of propagation and classically conserved quantities, much less to banish them from everyday discourse about everyday objects, but I think my argument does demand from any philosopher who wishes to invoke such a notion in his arguments an accounting of why he is justified in doing so, why his topic calls for that sort of notion, in light of the fact that there are strong grounds for believing such a notion cannot be, fundamentally speaking, true. In particular, any account of causality richer than the inital value formulation of mathematical physics that is supposed to arise naturally from an analysis of physical theory ought to be treated with suspicion.

Philosophers involved in projects ranging from arguments for the physical basis of the direction of time (*e.g.*, Reichenbach 1956), to the origin of linguistic reference (*e.g.*, Putnam 1975), to analysis of perception (*e.g.*, Russell 1927), to accounts of physical measurement (*e.g.*, Hacking 1983) and defenses of realism (*e.g.*, Boyd 1991, Hacking 1983 and Shimony 1993), no longer get access to such accounts of causality for free. That certain concepts do not accurately mirror the structure of the world at a fundamental level does not *eo ipso* preclude them from useful service in many areas of intellectual endeavor, but it does demand that such use be scrutinized. It would, for instance, be a strange (though possible) theory of linguistic reference that broke down in the vicinity of black holes—surely something would have to be said about why this ought to be so. Though this lesson about the circumscriptions on uses of certain causal notions perhaps could have been drawn from quantum mechanics alone,[42] such an argument would not have been nearly so clean and straightforward as that from general relativity, given the hotbed of dispute surrounding any interpretational theses forwarded about quantum phenomena.[43]

References

Arnowitt, R., S. Deser, and C. Misner (1962). The dynamics of general relativity. In L. Witten (Ed.), *Gravitation: An Introduction to Current Research*. New York: Wiley & Sons Press.

Bondi, H. (1962). On the physical characteristics of gravitational waves. In A. Lichnerowicz and A. Tonnelat (Eds.), *Les Théories Relativistes de la Gravitation*, Number 91 in Colloques Internationaux, pp. 129–135. Paris: Centre National de la Recherche Scientifique.

Bondi, H., M. van der Burg, and A. Metzner (1962). Gravitational waves in general relativity: VII. Waves from axi-symmetric isolated systems. *Philosophical Transactions of the Royal Society (London) A269*, 21–52.

[42] van Fraassen (1989) attempts a related project.

[43] I thank David Malament and Howard Stein for impeccable advising on and penetrating criticism of my doctoral thesis, from the third chapter of which this paper was harvested, and for many stimulating conversations on these and related topics. I also thank Robert Geroch for many stimulating, edifying conversations. If it were not too cheeky, I would thank him for being an unerring oracle on all topics physical and mathematical.

Boyd, R. (1983[1991]). On the current status of scientific realism. In *The Philosophy of Science*, pp. 195–222. Cambridge, MA: The MIT Press. Originally published in *Erkenntnis* 19(1983):45–90.

Bunge, M. (1979). *Causality and Modern Science* (third revised ed.). New York: Dover Publications, Inc.

Colombeau, J. (1992). *Multiplication of Distributions*. Number 1532 in Lecture Notes in Mathematics. Berlin: Springer.

Curiel, E. (1996). General relativity has no proper Hamiltonian formulation. Unpublished manuscript.

Curiel, E. (1999). The analysis of singular spacetimes. *Philosophy of Science 66*(Proceedings), 119–145.

Curiel, E. (2011). A formal model of the regime of a physical theory, with applications to problems in the initial-value formulation of the partial-differential equations of mathematical physics. Unpublished.

Ducasse, C. (1926). On the nature and the observability of the causal relation. In E. Sosa and M. Tooley (Eds.), *Causation*, Oxford Series in Philosophy, pp. 125–36. Oxford: Oxford University Press. Originally published in *Journal of Philosophy* 23(1926):57–68.

Earman, J. (1995). *Bangs, Crunches, Whimpers and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*. Oxford: Oxford University Press.

Eddington, A. (1923). *Mathematical Theory of Relativity* (second ed.). Cambridge: Cambridge University Press.

Einstein, A. (1916). The foundation of the general theory of relativity. In *The Principle of Relativity*, pp. 109–164. New York: Dover Press, 1952. Published originally as "Die Grundlage der allgemeinen Relativitätstheorie" *Annalen der Physik* 49(1916).

Einstein, A. (1922[1984]). *The Meaning of Relativity* (fifth ed.). Princeton: Princeton University Press. Trans. E. Adams, E. Straus and S. Bargmann. First edition published in 1922.

Geroch, R. (1973). Energy extraction. *Annals of the New York Academy of Sciences 224*, 108–117.

Geroch, R. (1996). Partial differential equations of physics. In G. Hall and J. Pulham (Eds.), *General Relativity*, Aberdeen, Scotland, pp. 19–60. Scottish Universities Summer School in Physics. Proceedings of the 46th Scottish Universities Summer School in Physics, Aberdeen, July 1995.

Hacking, I. (1983). *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

Hadamard, J. (1923). *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. New York: Dover Publications. A 1952 reprint of the original 1923 publication by Yale University Press.

Helmholtz, H. (1853). On the conservation of force; a physical memoir. In J. Tyndall and W. Francis (Eds.), *Scientific Memoirs from the Transactions of Foreign Academies of Science and Foreign Journals*, pp. 114–62. London: Taylor & Francis. Originally read before the Physical Society of Berlin (in German) on July 23, 1847. Translator unknown.

Lewis, D. (1986). Causal explanation. In *Philosophical Papers*, Volume 2, pp. 214–241. Oxford: Oxford University Press, 1986.

Mackie, J. (1980). *The Cement of the Universe* (paperback ed.). Oxford: Clarendon Press.

Malament, D. (1977). The class of continuous timelike curves determines the topology of spacetime. *Journal of Mathematical Physics 18*(7), 1399–***.

Maxwell, J. C. (1877). *Matter and Motion*. New York: Dover Publications, Inc. Originally published in 1877. This edition is a 1952 unaltered republication of the 1920 Larmor edition.

Mayer, J. (1842). On the forces of inorganic nature. In R. Lindsay (Ed.), *Energy: Historical Development of the Concept*, Volume 1, pp. 277–283. Stroudsberg, PA: Dowden, Hutchinson, Ross. Originally published as "Bemerkungen über die Kräfte der unbelebten Natur", *Annalen der Chemie und Pharmakie* 42(1842):233.

Mellor, D. (1995). *The Facts of Causation*. International Library of Philosophy. London: Routledge.

Mill, J. S. (1874). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation* (eighth ed.). New York: Harper & Brothers Publishers.

Norton, J. (1985). What was Einstein's principle of equivalence? *Studies in History and Philosophy of Science 16*, 203–246.

Norton, J. (1993). General covariance and the foundations of general relativity: Eight decades of dispute. *Reports on Progress in Physics 56*, 791–858.

Pauli, W. (1921). *The Theory of Relativity*. New York: Dover Publications, Inc. A 1981 reprint of the 1958 edition from Pergamon Press, a translation by G. Field of the original "Relativitätstheorie", in *Encyklopädie der matematischen Wissenschaften*, vol. V19, B. G. Teubner, Leipzig, 1921.

Peirce, C. S. (1898). *Reasoning and the Logic of Things*. Cambridge, MA: Harvard University Press. The Cambridge Conferences Lectures, delivered by Peirce between February and March of 1898, in Cambridge, MA. Edited by K. Ketner.

Penrose, R. (1968). Structure of spacetime. In C. DeWitt and J. Wheeler (Eds.), *Battelle Rencontres*. New York: W. A. Benjamin.

Putnam, H. (1975). Explanation and reference. In *Mind, Language and Reality: Philosophical Papers*, Volume 2, Chapter 11, pp. 196–214. Cambridge: Cambridge University Press. Originally published in *Conceptual Change*, eds. G. Pearce and P. Maynard (Dordrecht: Kluwer Academic Publishers, 1973), pp. 199–221.

Quine, W. (1973). *Roots of Reference*. LaSalle, IL: Open Court Press.

Quinn, T. and R. Wald (1999). Energy conservation for point particles undergoing radiation reaction. *arXiv:gr-qc/9903014v1* (http://xxx.lanl.gov/abs/gr-qc/9903014).

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press. Edited posthumously by Maria Reichenbach. Reprinted in 1991 with a foreward by H. Putnam.

Russell, B. (1927). *The Analysis of Matter*. New York: Dover Press. A 1954 reprint of the original 1927 edition.

Russell, B. (1948). *Human Knowledge: Its Scope and Limits*. London: George Allen and Unwin, Ltd. Third impression, printed in 1956.

Sachs, R. (1962). Gravitational waves in general relativity: VIII. The waves in asymptotically flat space-time. *Philosophical Transactions of the Royal Society (London) A270*, 103–126.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Shimony, A. (1993). Reality, causality and closing the circle. In *Search for a Naturalistic World View: Scientific Method and Epistemology*, Volume 1, Chapter 2, pp. 21–61. Cambridge: Cambridge University Press.

Sommerfeld, A. (1964). *Partial Differential Equations in Physics*, Volume VI of *Lectures on Theoretical Physics*. New York: Academic Press. Trans. E. Straus.

Stein, H. (1989). Yes, but. . . : Some skeptical remarks on realism and anti-realism. *Dialectica 43*, 47–65.

Stein, H. (1994). Some reflections on the structure of our knowledge in physics. In D. Prawitz, B. Skyrms, and D. Westerståhl (Eds.), *Logic, Metholodogy and Philosophy of Science*, Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy of Science, pp. 633–55. New York: Elsevier Science B.V. I do not have access to the published version of Stein's paper, but rather only to a typed manuscript. All references to page numbers, therefore, do not correspond to those of the published version. The typed manuscript I have is about 17 pages long, and the published version about 22. Multiplying the page numbers I give by $\frac{17}{22}$ and adding the result to 633 (the number of the first page in the published version) should give approximately the page number in the published version.

Stein, H. (unpub.). How does physics bear upon metaphysics; and why did Plato hold that philosophy cannot be written down? Typed manuscript not submitted for publication.

Tolman, R. (1934[1987]). *Relativity, Thermodynamics and Cosmology*. New York City: Dover Publications, Inc. A facsimile of the edition published by the Oxford University Press, at Oxford, 1934, as part of the International Series of Monographs on Physics.

van Fraassen, B. (1989). The Charybdis of realism: Epistemological implications of Bell's inequality. In J. Cushing and E. McMullin (Eds.), *Philosophical Consequences of Quantum Theory:*

Reflections on Bell's Theorem, pp. 97–113. Notre Dame, IN: University of Notre Dame Press. Originally published in *Synthese* 52(1982):25–38. Appendix added in 1989.

Vickers, J. and J. Wilson (1998). A nonlinear theory of tensor distributions. *arXiv:gr-qc/9807068v1* (http://xxx.lanl.gov/abs/gr-qc/9807068).

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

Wald, R. (1999). Gravitation, thermodynamics and quantum theory. *arXiv:gr-qc/9901033v1* (http://xxx.lanl.gov/abs/gr-qc/9901033).