December 21, 2007; January 21, 2008

Is There an Independent Principle of Causality in Physics? A Comment on Matthias Frisch, "Causal Reasoning in Physics."

John D. Norton
Department of History and Philosophy of Science
Center for Philosophy of Science
University of Pittsburgh
www.pitt.edu/~jdnorton

Matthias Frisch has argued that the requirement that electromagnetic dispersion processes are causal adds empirical content not found in electrodynamic theory. I urge that this attempt to reconstitute a local principle of causality in physics fails. An independent principle is not needed to recover the results of dispersion theory. The use of "causality conditions" prove to be either an exercise in relabeling an already presumed fact; or, if one seeks a broader, independently formulated grounding for the conditions, that grounding either fails or dissolves into vagueness and ambiguity, as has traditionally been the fate of candidate principles of causality.

1. Introduction

In his (manuscript a) and revision (manuscript b) that remarks[1] on an earlier draft of this note, Matthias Frisch responds to a skeptical tradition to which I have contributed (Norton 2003, 2007). That tradition doubts that the sciences are founded upon an independent principle of causality. His response leads him to argue (manuscript a, b, Section 4) that the requirement that

[1] I thank Matthias for these remarks, which have helped me correct and clarify the points I make here.

certain physical processes are causal does add further physical content. His example is dispersion in classical electrodynamics. There causal considerations are invoked at a decisive moment in the derivation of the dispersion relations, purportedly to provide physical content not recoverable through the usual manipulations of electodynamic theory.

In this sense, Frisch is proposing a principle of causality. His principle asserts that effects cannot precede their causes. It is independent in the sense that it provides factual content not supplied by the relevant physical laws. Since he is sure only that it applies to a few physical processes, Frisch allows that the principle may not hold universally. So it is the proposal for a local, independent principle of causality.

While Frisch's example is both important and intriguing, my purpose in this note is to argue that Frisch is mistaken in his analysis of it. After a brief review of dispersion theory in Section 2, I will urge in Section 3 that the causal constraints at issue are merely shorthand for physical constraints already recoverable in classical electrodynamics, though possibly not easily recoverable. While I believe that we need summon no causal metaphysics to complete dispersion theory, in Section 4, I will explore the consequences of persisting in efforts to do just this. Those efforts lead in two directions. In one, we merely end up assigning an additional adjective "causal" to a condition we believe on other grounds. In the other, we seek a precise, independent expression, usable in physical theorizing, for the general requirement that effects cannot precede their causes in processes like dispersion. Efforts to formulate this principle independently lead to failure or vagueness and ambiguity.

2. Scattering in Classical Electrodynamics

In classical electrodynamics, a dielectric scatters an incoming field. In roughest outline, the basic supposition is that the scattered field at a point **x** in space in the dielectric at time t, written here "*scattered* (**x**,t)," depends linearly on the incident field "*incident* (**x**, t')" at the same point **x** and other times t'. Hence *scattered* (**x**,t) can be reconstructed if we know just which scattered fields G(**x**, t) arise from a delta function incident field, at the same point **x** but massed at time 0. Linearity allows us to recover the scattered field from an arbitrary incoming field by the integration

$$scattered(\mathbf{x},t) = \int_{-\infty}^{\infty} G(\mathbf{x},t')\,incident(\mathbf{x},t-t')\,dt' \tag{1}$$

In this integral, the scattered field at **x** at time t is computed as a weighted sum of the incident field values at that same point **x** but at different times. The quantity G(**x**,t') determines the times for which the incident field at **x** contributes. The causality condition at issue requires that no incident field at a time later than t can contribute to the scattered field at t. It is enforced by requiring that[2]

$$G(\mathbf{x},t') = 0, \text{ for all } t'<0 \qquad (2)$$

for, with this condition, *incident*(**x**, t-t') can make no contribution to the integral of (1) whenever t-t'>t, that is, whenever t'<0.

3. Sufficiency of the Physics

The question at issue is the physical foundation of this condition (2). Frisch believes that it is founded upon a principle of causality that is independent of electrodynamic theory in the sense that it places additional factual constraints on the theory. I maintain that the condition can and should be founded upon existing electrodynamic theory alone.

To see why I hold this latter view, recall the physical process at issue. A dielectric in classical theory consists of bound charges. They are initially at rest. An incident field impinges upon them. When that field arrives or perhaps when an especially strong peak in that field arrives at the position **x**, only then does it accelerate the charge at **x** by virtue of the Lorentz force law. Subsequently, by virtue of Maxwell's equations, the accelerated charge at **x** will radiate the scattered field.

In this computation, the causality condition is a consequence of the laws of electrodynamics and our decision to describe a scattering process. The laws entail that no electromagnetic process can propagate at faster than c, the speed of light. These laws permit the time reversal of scattering processes (discussed more in Section 4.2 below). In this time reversed process, the present state of the system depends only upon its future states—that is just the time reversal of the condition applying in ordinary scattering that the present state of the system depends only upon its past states. We preclude the time reversed process by stipulating that we wish to describe the ordinary scattering process and that stipulation enters into the computation through suitable boundary conditions.

---

[2] Or that the integral of (1) be computed only between the limits of t'=0 and t'=∞.

Frisch quotes Jackson's (1999) text as his primary source. Yet Jackson (in my edition) does not explicitly deduce (2) from an independent principle of causality or give a precise formulation of such a principle. What he does do, as Frisch notes, is to deduce the condition (2) from standard electrodynamics for a special case (Section 7.10.B) without drawing on causality conditions. He then observes (Section 7.10.C) that this outcome is "in accord with our fundamental ideas of causality in physical phenomena" and finally announces that (1) and (2) combined comprise "the most general spatially local, linear, and causal relation…"

The development does not make clear how the step from the special to the general case is taken. It might be, as Frisch suggests, that Jackson is calling upon some more fundamental principle of causality that lies outside classical electrodynamics. Or it might just be that Jackson is following a more benign approach that needs no additional physical principles. That is, he is suggesting that, in all more complicated cases, a more detailed and possibly very difficult analysis fully within electrodynamics would return the same result (2). Since he not giving the analysis but nonetheless is confident of its outcome, it is an awkward point to make. So he tries to make the result plausible by noting that it fits with causal expectations, although a precise causal principle is not actually formulated and used to deduce the result.[3]

4. Failure of the Principle of Causality Proposed

While I believe a principle of causality is not needed to complete dispersion theory, we can ask what are our prospects of formulating a non-trivial principle that could serve this

---

[3] There is a similar ambiguity in another source Frisch cites. Toll's (1956) principal goal is to prove the logical equivalence of dispersion relations in a fairly general context and "strict causality." The latter asserts that "no signal can travel faster than [c]" and is also glossed by a condition that is not obviously equivalent to it, "no output can occur before the input." (p. 1760) However Toll does not make clear whether strict causality is a universal principle to be required independently of all physical theories or merely a result to be discovered and demonstrated within each physical theory under consideration. Remarks in the concluding section suggest the latter. For Toll finds it (p. 1770) "an open question whether or not strict causality is a valid physical hypothesis" and then considers merely as a possibility that "strict causality … prove[s] to be invalid or unenforceable in future theories."

purpose. The question is worth exploring since Frisch at least interprets the physics literature as asserting that the derivation of results in dispersion theory are completed by invoking an independent principle of causality.

Such attempts require a great deal more than a vague gesture at "our fundamental ideas of causality." They require the presentation of a viable principle of causality in a sufficiently precise form for its applicability and proper functioning in this case to be apparent. My contention in this section is that neither Jackson nor Frisch formulates such a principle. Frisch's candidate for the applicable principle is that (Section 4.)

"effects do not precede their causes"

I will seek to show the following. When we try to formulate a general statement of the principle in terms sufficiently precise for physical theorizing, the principle either fails or becomes too vague to use. If, however, we retract and merely declare the one case of dispersion to be an implementation of a principle for which we offer no more general statement, then all we have achieved is a relabeling of the one case in suggestive but physically empty causal language.

4.1 A Sometimes Principle

Frisch does not want to commit to the universal applicability of this principle. "It might in fact be true," he allows, "…But I think we should allow for the possibility that a certain causal principle or assumption is not true in general and nevertheless take it to be physically well founded, in a certain sense." (Section 4)

If Frisch is serious about this possibility, then his sometimes principle may be no independent principle at all. It looks like a principle that holds, except when it doesn't. For, speaking figuratively, how are we to know whether it applies to some system?[4] We must call

---

[4] Frisch (manuscript b, p.17, fn. 8; see also p.18) responds by affirming that we know scattering processes are causal in the same way that we know that any theoretical property holds: we find its presumption leads to successful predictions. As revisions to this text now make clear, this response conflates epistemic and figurative sense of "know." Frisch has remarked on how we learn the property obtains for some system. My challenge is not epistemic, but concerns the condition under which the principle of causality applies to a system. What properties must it

upon the properties of that system and affirm that the principle holds for them. That is, the obtaining of the principle of causality threatens merely to be a restating of the properties of the system already known. The danger is that the conformity of the system to the principle places no additional factual constraint upon it. It that sense, it would be a mere honorific.

That threat is realized if we retract to the safety of merely considering dispersion processes in isolation. For then we can comfortably declare the incident field a "cause," the scattered field an "effect" and the relation (2) of dispersion theory as expressing the principle of causality in that it assures us that these particular effects never precede their causes. If we just consider dispersion theory in isolation, the exercise is purely one of labeling. Nothing is added by the causal talk beyond restating the specific result already at hand in new causal language.

That exercise in labeling is clearly not what Frisch intends. But if the exercise is to be anything more, there must be some relation to systems and processes outside dispersion theory. Of course there is such a connection in an informal sense. The incident wave is analogous to my shout and the scattered wave to the startled cats, fleeing from the overturned cream jug. But that informal connection merely makes the application of the causal language comfortable. It does not locate dispersion processes within broader factual regularities; and it does not supply a theoretical instrument of sufficient precision to enable completion of physical computations in dispersion theory. What we need is some more general property of the system that would mark it antecedently as causal. Then we would know antecedently that the local principle of causality must apply to it and that it falls into a greater causal order in nature, even if not a universal order.

In the following I will investigate what is needed to provide a more general principle that is also precise enough for application in dispersion theory and other physical theories.

4.2 The Conditions of Applicability are Obscure

Let us presume that Frisch's principle of causality is formulated precisely enough for us to apply it in physical theories. If the principle is to serve as indicated in scattering theory, it must be clear that the principle applies. To see that its applicability is questionable, we need to recall that classical electrodynamics is a time reversible theory. If the theory allows a process,

---

have if the principle is to apply; that is, figuratively, if we know the system has those properties, then we know that principle applies?

then it also allows its time reverse. The theory allows a dielectric to scatter an incident wave. Therefore it also allows a time reversed, scattered wave to collapse back onto the dielectric and return the time reverse of the incident wave. Of course this reverse process is highly unlikely in ordinary circumstances, just as it is possible, but highly unlikely, for ripples in a pond to converge and eject a stone.

Now imagine a universe completely empty excepting two processes that we will call “A” and “B”. Process A has an incident wave, a dielectric and a scattered wave. Process B is the time reverse of A. The two processes are completely isomorphic in all properties. Any property of one will have its isomorphic correlate in the other. Any fact about one will have a correlate fact obtaining for the other. One might be tempted to imagine that one of the two processes is “really” the ordinary one, progressing normally in time; while the other is a theoretician’s fantasy, a possibility in principle, but in practice unrealizable. The essential point of the example is that no property of the A and B systems distinguish which is which. Every property of one has a perfect correlate in the other.

Let us assume that Frisch’s principle of causality applies to one of these processes, the A process, for example. That will be expressed as a condition that the present state of the process depends only on its past states. Exactly what “depends” may amount to is to be decided by the principle. All that matters for our purposes is that an exactly isomorphic condition of dependence will obtain in the B process, except it will be time reversed. Indeed using the time order natural to process A, we would have to say that the principle of causality requires the present states of process B to depend upon its future states.

In short, if the principle applies to process A, it fails for process B; and conversely. This is a reductio ad absurdum of the applicability of Frisch’s principle of causality to scattering in classical electrodynamics. Or, to put the outcome another way, if the principle holds for one process but not the other, then the decision as to which process is properly causal cannot be based in any physical difference between the two processes. For every physical property of one process has an exact correlate in the other. The declaration that the principle holds for one process but not the other has become an arbitrary stipulation without physical basis.

It seems to me that there is only one escape. It is to propose that there is, as a factual matter outside electrodynamics, a natural time direction. When we require that electrodynamics must respect that direction, we are able to preclude one of the two processes. Possibilities of this sort have been repeatedly offered, weighed and found wanting. Nearly a century ago, for example, Ritz (in Einstein and Ritz, 1909) urged that electrodynamics should be formulated in terms of retarded potentials only, thereby denying the time reversibility of the theory. The mainstream agreed with Einstein's response. He insisted that the time reversed processes were possible, just statistically very unlikely.[5]

To get a sense of why the mainstream has flowed in this direction, recall that the time reversed process can be broken up into many small parts. Locally, each small section of the collapsing incident wave is merely a wavefront propagating in quite ordinary ways; momentarily, the force exerted by the incident wave on the unscattering charges just follows the Lorentz force law; and so on throughout. It is only when the many small pieces are assembled that we find an unfamiliar process. It seems mistaken to invoke additional laws to prohibit a total system, perfectly admissible in all its parts. Indeed we have little doubt that were we somehow to contrive a perfect time reversal of a scattered wave, its future course would be the time reverse of ordinary scattering.

Of course Frisch is well aware of the problems of combining a time asymmetric principle of causality with a time symmetric physical theory and discusses them at some length in his Section 3. None of the arguments given there escape the difficulty just described for his principle of causality. His most interesting proposal is that we might recover the time asymmetry of causation empirically. We intervene in a process and discover as a matter of experimental fact that our intervention perturbs the future and not the past.

It is precisely to avoid such considerations that my example of the A and B processes assumes an otherwise empty universe in which there are no agents to intervene. In any case, it is unclear what the intervention experiments reveal. The system at issue is now exceedingly complicated and poorly understood: it is the scattering system plus human beings who poke their fingers into the beams and have sensors—eyes—presumed not to emit radiation but only to absorb it. How are we to know whether asymmetries arising in a system that complicated are due

---

[5] Ritz's proposal is of no help to Frisch's principle of causality. For with it, no principle of causality is needed to complete the scattering computation. The restriction to retarded potentials is a precise electrodynamic expression of the restriction that the present electrodynamic state depends only on it past state.

to an intrinsically asymmetric causal relation? Or are they due to some asymmetry introduced into an otherwise fully time symmetric theory through the conditions that describe the vastly complicated system of humans with intervening fingers and watching eyes? And do we doubt that, if those fingers could bring about the time reversal of a scattered wave as an initial condition, then the time reverse of scattering would ensue?

4.3 Effects Can Come Before Their Causes

What of the principal content of the principle, that effects cannot precede causes? Frisch supports his principle of causality with the remark neither of my earlier writings (Norton 2003, 2007) includes counterexamples to it. That lacuna is easily remedied here, for there are many cases in which the effect preceding the cause is accepted as a possibility.

There is a flourishing literature on the physical possibility of time travel. (Arntzenius and Maudlin, 2005) It is no longer believed that problems of causality—the "grandfather paradox"—preclude its physical possibility. Every instance of time travel involves effects preceding their causes.

Another example arises in the case of tachyons in special relativity. If one observer judges a tachyon to be propagating forward in time, then we can always find a second observer who is moving inertially with respect to the first and who judges the time order of the states to be reversed. If the first observer judges that the direction of propagation coincides with the direction of causal action, then the second observer would describe that judgment as affirming that the later states causally affect the earlier states.[6]

Finally there is a third example in the A and B processes above. Let us say that there is some sense in which the incident wave of A is the cause of the scattered wave of A, the effect. Then under the isomorphism indicated above, we will conclude that the later state of the time reversed incident wave of B is the cause of the earlier time reversed scattered wave of B.

---

[6] Should we judge tachyons physically impossible since this sort of backwards causation might lead to paradoxes? While that concern has been debated at some length in the literature, I agree with Artnzenius' (1990) conclusion that paradox-free tachyonic theories are possible.

The idea that effects might precede their causes has sufficient currency that it has become the subject of an article, "Backwards Causation" in the *Stanford Encyclopedia of Philosophy*. See Faye (2005).

4.4 Vagueness of the Relata and of the Notion of Causal Process

If Frisch's principle of causality is to be applicable to physical systems, we must know *precisely* what is a cause; what is an effect; and which processes are causal. Otherwise we will not know how to apply the principle in concrete cases of physical theorizing.

Consider the first two questions. What precisely is a cause? What precisely is an effect? Can anything be a cause and an effect? Surely there must be some restrictions and they need to be made explicit. Presumably, for example, causes and effect must be the states of physical systems restricted in time, for the principle makes assertions on their time order. Are they any state extended over some short time interval? Or are they any state at an instant? If we are in special relativity, which instant do we choose? The relativity of simultaneity allows many competing instants according to the inertial frame of reference chosen. If we seek to avoid the problem by requiring the states to be localized at a point in space, are we thereby precluding by fiat that non-local quantum states can enter into causal relations?

Now the third question: What sorts of processes are properly labeled as causal? In the case of scattering, causal dependence was translated into the mathematical dependence of a field at one time on another field at other times. Are all such dependencies causal processes? If so, what are we to make of Lagrange principles? A particle's trajectory is picked out as the one that extremizes the integrated action for its entire history between a start and an end point. As a result, the computation of the acceleration of the particle now involves its motion at earlier and later times. Is this dependence causal? If so, we have a violation of Frisch's principle of causality, for a future state is affecting the present state. If it is not causal, how are we to pick out the computations that correspond to real causal processes?

Finally, when we have a causal process, how do we decide which of the relata is the cause and which is the effect? We cannot merely use their time order to decide; that is, we cannot just pick out the cause as the earlier relata and the effect as the later. That reduces the principle that causes cannot precede their effects to a definition. If the principle is to have physical content, it must provide an independent characterization of what intrinsically distinguishes a cause from an effect. What is that characterization?[7]

5. Conclusion

In sum, it is not so hard to find vague and sometimes grand sounding causal talk in the physics literature. It is easy to yield to the temptation of saying that this shows that there really is an independent principle of causality at work in physics. There is, however, a chasm between vague and grand causal talk and a precise principle of causality that can and should be used to augment the physical content of existing physical theories. Frisch's proposal has not closed that gap and, on the basis of the many obstacles just presented, it seems unlikely that this gap can be closed.

In Norton (2003), I suggested that causal fundamentalism and its associated principle of causality succumbed to a dilemma: either the view places factual constraints on the world (in which case they proved false); or they did not (in which case the causal labels were honorifics). Frisch's attempt to locate physical content in the idea that dispersion is causal meets an analogous fate. In so far as it is precisely stated, his causality condition is not derived from a broader regularity; it merely attaches the adjective "causal" to a particular result in scattering theory. When we seek the broader regularity, we find that it most likely cannot be stated precisely enough for application in physical theorizing, even allowing that this broader regularity need not hold universally.

References

Arntzenius, Frank (1990), "Causal Paradoxes in Special Relativity," *The British Journal for the Philosophy of Science*, **41**, pp. 223-243.

---

[7] One might call upon Reichenbach's famous idea that real causal processes are distinguished as those that can carry a "mark." That idea is of no help to Frisch's principle of causality since, as Grünbaum (1973) has argued, the mark criterion fails to pick out a direction in time and reveals only causal relatedness.

Arntzenius, Frank and Tim Maudlin, (2005) "Time Travel and Modern Physics", *The Stanford Encyclopedia of Philosophy (Summer 2005 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/sum2005/entries/time-travel-phys/>.

Einstein, Albert and Ritz, Walter, (1909) "Zum gegenwärtigen Stand des Strahlungsproblems" *Physikalische Zeitschrift,* **10**, p. 554.

Faye, Jan (2005) "Backward Causation", *The Stanford Encyclopedia of Philosophy* (Fall 2005 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/fall2005/entries/causation-backwards/>.

Frisch, Matthias (manuscript a) “Causal Reasoning in Physics” In [2007] EPSA07: 1st Conference of the European Philosophy of Science Association (Madrid, 15-17 November, 2007), philsci-archive.pitt.edu/archive/00003732

Frisch, Matthias (manuscript b) “Causal Reasoning in Physics” In [2008] Causation Workshop (Pittsburgh, PA; January 26, 2008), philsci-archive.pitt.edu/archive/00003820

Grünbaum, Adolf (1973) “The Causal Theory of Time,” pp. 180-208 in *Philosophical Problems of Space and Time: Boston Studies in Philosophy of Science, Vol. XII*. R. S. Cohen and M. W. Wartofsky, eds., Dordrecht: Reidel.

Norton, John D. (2003) "Causation as Folk Science," *Philosophers' Imprint* Vol. 3, No. 4 http://www.philosophersimprint.org/003004/; reprinted as pp. 11-44 in H. Price and R. Corry, *Causation and the Constitution of Reality: Russell’s Republic Revisited*. Oxford: Clarendon Press.

Norton, John D. (2007) "Do the Causal Principles of Modern Physics Contradict Causal Anti-Fundamentalism?" pp. 222-34 in Thinking about Causes: From Greek Philosophy to Modern Physics. eds. P. K. Machamer and G. Wolters, Pittsburgh: University of Pittsburgh Press.

Jackson, John David (1999) *Classical Electrodynamics*. 3rd ed. New York: Wiley, 1999.

Toll, John S. (1956) “Causality and Dispersion Relation: Logical Foundations,” *Physical Review*, **104**, pp. 1760-70.