

No superluminal propagation for classical relativistic and relativistic quantum fields

John Earman

Dept. of History and Philosophy of Science

University of Pittsburgh

Pittsburgh, PA 15217, USA

ABSTRACT: A criterion is proposed to ensure that classical relativistic fields do not propagate superluminally. If this criterion does indeed serve as a sufficient condition for no superluminal propagation it follows that various other criteria found in the physics literature cannot serve as necessary conditions since they can fail although the proffered condition holds. The rejected criteria rely on energy conditions that are believed to hold for most classical fields used in actual applications. But these energy conditions are known to fail at small scales for quantum fields. It is argued that such a failure is not necessarily a cause for concern about superluminal propagation in the quantum regime since the proffered criterion of no superluminal propagation for classical fields has a natural analog for quantum fields and, further, this quantum analog condition provably holds for some quantum fields despite the violation of energy conditions. The apparatus developed here also offers a different approach to treating the Reichenbach-Salmon cases of “pseudo-causal processes” and helps to clarify the issue of whether relativity theory is consistent with superluminal propagation.

Keywords: relativity theory, causal propagation, causal processes vs. pseudo-causal processes, relativistic energy conditions

E-mail address: jearman@pitt.edu

1. Introduction

Various energy conditions have been formulated for classical relativistic fields, including the so-called dominant energy condition which has been

interpreted as saying that matter-energy does not flow faster than light. It will be argued here that this condition is not necessary for no superluminal propagation. The argument takes the form of formulating and defending a sufficient condition for no superluminal propagation and of showing that this sufficient condition can hold even when various energy conditions fail, including the condition of no negative energy densities (which is entailed by the dominant energy condition).

This conclusion has little impact in actual physical applications covered by classical field theories because most of the known fields satisfy the energy conditions in question. By contrast quantum fields generically violate the classical energy conditions; in particular, the energy density at a spacetime point can be unboundedly negative. It is shown that the algebraic approach to quantum field theory permits the formulation of a natural analog of the proffered sufficient condition for no superluminal propagation for classical fields. It is argued that in models of QFT where this analog condition holds there should be no worries about superluminal propagation even at the small scales at which there are violations of energy conditions.

The approach developed here also suggests a new way of addressing one of the main issues discussed in the philosophical literature on “causal processes”; namely, how to recognize and deal with examples of pseudo-causal processes where something is moving faster than light. And, following Geroch (2011), this approach also has implications for the issues of whether relativity theory is consistent with superluminal processes and whether such processes would produce intolerable “paradoxes.”

2. (NSP1) Initial value formulation

To secure one sense and, it will be argued, the most fundamental sense of no superluminal propagation for classical relativistic fields, one looks to an initial value formulation for the equations governing the fields. The sought after formulation has two parts, an existence part and a uniqueness part, that have respectively the following schematic forms:

For any initial value hypersurface S and any initial datum Φ_0 on S

- 1) there is an open neighborhood U of S and a solution Φ of the field equations on U that agrees with Φ_0 , and

2) for any point $p \in U$ if p belongs to the domain of dependence $D(A)$ of a closed subset A of S , then for any solutions Φ and Φ' on U that agree with Φ_0 on A , $\Phi(p) = \Phi'(p)$.¹

In standard relativity texts S is taken to be a hypersurface that is spacelike with respect to the spacetime metric and the domain of dependence $D(A)$ of $A \subset S$ is defined to be the set of all of those spacetime points q such that any curve γ in U meets A in exactly one point if γ passes through q and if the tangent to γ at any of its points is contained in the null cone of the spacetime metric at that point. An initial value formulation with this definition of domain of dependence is denoted by (IVF).

The local nature of the existence part of the initial value formulation is undesirable since what is wanted is a global existence result. But Nature does not always oblige since a solution may break down after a finite time due to the development of singularities. This problem is sometimes pushed aside in the context of special relativistic theories by dismissing those fields for which there is a failure of global existence of solutions as failing to be “fundamental fields.”² The problem cannot be pushed aside in general relativistic setting where singularities are a generic feature of solutions to Einstein’s gravitational field equations. I will ignore this topic here to concentrate on the uniqueness part of the initial value formulation.

In contrast to the existence part, the local nature of the uniqueness result is highly desirable; indeed, I propose to gloss it as saying that the fulfillment of (IVF) implies that the field propagates non-superluminally, and I will use (NSP1) to stand for this proffered sufficient condition for non-superluminal propagation. The justification for thinking that (NSP1) does indeed serve as a sufficient condition will be discussed below. In advance of that discussion, note that (NSP1) is equivalent to another natural candidate for no superluminal propagation of the field; namely, changes in the initial data on an arbitrary closed subset A of the initial data surface, while holding fixed the initial data in the rest of S , does not result in changes in the solutions to the field equations outside the region $J(A)$ of the causal influence of A .³ ($J(A)$

¹For present purposes I leave aside the issue of continuous dependence on initial data, which is taken to be part of a “well-posed” initial value problem.

²This dismissal seems justified in instances where the “blow up” of the solution is an artifact of an idealization, as with dust matter.

³The equivalence follows from the fact that $D(A)$ is the spacetime complement of $J(S/A)$. Note that if the field equations entail constraint equations—as with Maxwell’s

consists of all spacetime points such that there is a curve γ from A to p such that the tangent to γ at any of its points is contained in the null cone of the spacetime metric at that point.) Of course, if one thinks that causal influence only “goes” in the future direction, then one will want to focus on the causal future, $J^+(A)$, of A . (Introducing a time orientation, $J^+(A)$ consists of all points p such that there is a future directed causal curve from A to p .) Thus, Wald (1984, p. 244) writes: “[C]hanges in the initial data in a region, A , of the initial data surface should not produce any changes outside the causal future, $J^+(A)$, of this region. If such changes occurred, we should be able to use them to propagate signals ‘faster than light’.”⁴

Geroch (2011) has noted that the point of view just sketched is parochial in a way that obscures important foundational issues.⁵ An initial value formulation of the above form cannot be expected unless the field equations can be massaged to take the form of a system of symmetric hyperbolic partial differential equations (pdes).⁶ But such a system defines its own causal cones with respect to which an initial value formulation holds when the domain of dependence is defined in terms of these cones.⁷ I will denote an initial value formulation with this broader understanding of domain of dependence by (IVF*). In some familiar cases, e.g. a massive Klein-Gordon field or a Maxwell electromagnetic field propagating on Minkowski spacetime or, more generally, a minimally coupled Klein-Gordon field or Maxwell field propagating on a curved Lorentzian spacetime, the causal cones defined by the relevant system of hyperbolic pdes coincide with the null cones of the spacetime metric and, thus, (IVF*) and (IVF) are equivalent. Indeed, it is such coincidences that endow the null cones of the spacetime metric with causal significance, and in the case of the Maxwell field the coincidence justifies giving the null cones the alternative appellation of “light cones.”

However, the causal cones defined by the hyperbolic system pdes governing some of the fields need not coincide with the null cones of the spacetime

equations for the electromagnetic field and Einstein’s field equations for the gravitational field—then the considered changes in the initial data in the region A may be limited by the condition that the data in S/A are held fixed.

⁴I have changed Wald’s notion to conform to that used here.

⁵In a sense, the present paper is just a long footnote to Geroch’s paper. But unlike many footnotes, I think this one is worth the trouble of reading it.

⁶For definitions of the relevant concepts and a precise treatment of the initial value problems for such pdes, see Geroch (1996) and Randall (2008).

⁷In general the causal cone at a point depends on the field values at that point.

metric—in particular, the former can be “wider” than the latter. Because they admit an initial value formulation (IVF*) the evolution of such fields is properly deemed to be causal even though it is superluminal. Worries about whether superluminal propagation will lead to paradoxes in the form of inconsistent causal loops need to be addressed. This matter will be taken up below in Section 9. In the meantime I concentrate on cases where the causal cones coincide with the null cones of the spacetime metric and, thus, the narrower form (IVF) of (IVF*) holds.

3. (NSP2) Causal energy-momentum flow

When concerns about superluminal propagation are raised they focus, of course, on superluminal flow of mass-energy. Thus, it might seem that the best way to address the core concern is not to proceed indirectly through the initial value problem but rather directly through the stress-energy-momentum tensor (which is usually shortened to stress-energy tensor). I do not tackle here the interesting question of how a stress-energy tensor T^{ab} is associated with a field but simply quote the standard examples from the physics literature. Given such a T^{ab} the energy-momentum density ${}_V P^a$ as measured by an observer with future-directed timelike four-velocity V^b is defined to be ${}_V P^a := -T^{ab}V_b$.⁸ The requirement that ${}_V P^a$ is non-spacelike for *all* observers is known as the dominant energy condition (DEC). It is equivalent to the statement that $T^{ab}V_aW_b \geq 0$ for all pairs V^a and W^b of future-directed timelike vectors. DEC entails the weak energy condition (WEC) which requires that the energy-density as measured by any observer is non-negative, i.e. $T^{ab}V_aV_b \geq 0$ for any timelike V^b .

Here are some typical glosses on DEC from relativity textbooks: the DEC “can be interpreted as saying that the speed of energy flow of matter is always less than [or equal to] the speed of light” (Wald 1984, p. 219); DEC has been given the interpretation that “energy cannot flow more than the speed of light” (Rendall 2008, p. 31). And an authoritative review article for the *Oxford handbook of the philosophy of physics* opines that DEC “can be interpreted as the requirement that matter cannot travel faster than light” (Manchak 2013, p. 595). Thus, a second plausible candidate (NSP2)

⁸The minus sign is needed because of the choice of the signature (+ + +−) of the spacetime metric.

criterion for no superluminal propagation is the requirement that the stress-energy tensor of the field satisfy DEC.

Illustrations of stress-energy tensors that satisfy DEC and some that violate it will be given below. But first I turn to a third attempt to specify the conditions for no superluminal propagation.

4. (NSP3) Causal propagation of matter

Referring to the initial value formulation, let Φ be a solution to the field equations on an open neighborhood U of the initial value hypersurface S . No superluminal propagation of matter has been formulated as the requirement (NSP3) that whenever the stress-energy tensor T^{ab} of the field vanishes on $A \subset S$ it is also vanishes throughout $D(A)$.

(NSP3) is closely related to DEC. Indeed, Hawking and Ellis (1972, lemma 4.3.1) show that together DEC and the condition $\nabla_a T^{ab} = 0$ entail (NSP3).⁹ Hawking and Ellis opine that this result “may be interpreted as saying that the dominant energy condition [with the help of $\nabla_a T^{ab} = 0$] implies that matter cannot travel faster than light” (1972, p. 94). The vanishing of the covariant divergence of the stress-energy tensor is sometimes referred to as the conservation condition, although it is only in a restricted class of relativistic spacetimes that this condition can be integrated to give a familiar form of the conservation of energy, saying that the energies of a system at any two times differs by the amount of energy-momentum that flows through the boundaries of the system in between those times.¹⁰

The relationships among the three conditions (NSP1)-(NSP3) for no superluminal propagation are complicated. What I want to focus on is the fact that (NSP1)—which I take to be the fundamental expression of no superluminal propagation—can hold while (NSP2) and (NSP3) both fail. I give some concrete examples to illustrate this fact and then turn to a discussion of its implication for the notion of causal propagation.

5. (NSP1) without (NSP2) or (NSP3)

⁹Here ∇_a is the unique covariant derivative compatible with the spacetime metric.

¹⁰A necessary and sufficient condition is that the spacetime admit a timelike vector field ξ^a such that $\nabla_{(a}\xi_{b)} = 0$. Minkowski spacetime, of course, satisfies this condition. Defining the energy-momentum (density) four-vector by $\xi P^a = -T^{ab}\xi_b$ it follows that $\nabla_a(\xi P^a) = 0$. The four-dimensional version of Gauss’ theorem can then be applied to derive the desired form of conservation of energy.

Ex. 1. Perfect fluid. The stress-energy tensor is

$$T^{ab} = (\mu + p)W^aW^b + pg^{ab}$$

where μ is the mass-energy density of the fluid, p is the pressure, W^a is the four-velocity of the fluid, and g_{ab} is the spacetime metric.¹¹ It is required that $\mu \geq 0$ so that WEC is satisfied and also that W^a is normalized (i.e. $W^aW_a = -1$).¹² The DEC applied to a perfect fluid entails that $\mu \geq |p|$. The conservation condition $\nabla_a T^{ab} = 0$ is equivalent to the Euler equations. When the equation of state $p = F(\mu)$ is such that $dF/d\mu > 0$ the Euler equations can be written in symmetric hyperbolic form and, thus, possess an initial value formulation of the desired form.¹³

Furthermore, when $1 > dF/d\mu$ the velocity of sound v_s as given by

$$\left(\frac{v_s}{c}\right)^2 = \frac{dF}{d\mu}$$

is less than the velocity of light c ($\equiv 1$) and the “sound cones” contain only timelike vectors so that (NSP1) holds for the perfect fluid. When $1 < dF/d\mu$ the velocity of light is exceeded by the velocity of sound and (NSP1) fails. Nevertheless, there is still a good initial value formulation in the sense that (IVF*) holds (see Geroch 2011). But for the present I stick to the case where the velocity of sound does not exceed the velocity of light. When $\mu \geq |p|$ (NSP2) and (NSP3) both hold alongside (NSP1).

But consistent with $1 > dF/d\mu > 0$ and, thus, with (NSP1), there can be a violation of DEC. For example, with $F(\mu) = 2 + \frac{1}{2}\mu$ DEC fails when $0 < \mu < 4$, $2 \leq p < 4$.¹⁴ In this regime (NSP2) and (NSP3) both fail while (NSP1) holds.

¹¹In the present examples one can keep things simple by assuming that the matter fields are propagating on a fixed background spacetime with metric g_{ab} . If g_{ab} is allowed to be a dynamical field then its field equations—say, Einstein’s gravitational field equations—have to be supplied, and then attention shifts to the coupled matter-Einstein equations.

¹²The latter is an example of a constraint equation. This aspect of the initial value problem was suppressed in the above presentation. For a good IVF it has to be shown that the constraints propagate, i.e. if they are satisfied at the initial time then they are satisfied at later times.

¹³The constraint $W^aW_a = -1$ does propagate.

¹⁴Such an equation of state does not describe actual fluids for which one expects the pressure to go to zero when the density of the fluid goes to zero. But this does not affect the conceptual point being made here.

Ex. 2. Linear scalar field. The stress-energy tensor for the scalar field ϕ is

$$T^{ab}(\phi) = \nabla^a \phi \nabla^b \phi - \frac{1}{2} g^{ab} (\nabla_c \phi \nabla^c \phi + V(\phi))$$

where $V(\phi)$ is the potential of the field. The equation of motion

$$\nabla^a \nabla_a \phi - \frac{dV}{d\phi} = 0$$

follows from Lagrange's equation for the Lagrangian $L = -\frac{1}{2} \nabla_c \phi \nabla^c \phi - V(\phi)$.¹⁵ In turn the equation of motion entails the local conservation law $\nabla_a T^{ab}(\phi) = 0$. For $V(\phi) = m^2 \phi^2$ we get the Klein-Gordon equation for a field of mass m

$$\nabla^a \nabla_a \phi - m^2 \phi = 0 \tag{KG}$$

(KG) is already in symmetric hyperbolic form so we know immediately that there is a good initial value formulation, and thus (NSP1) holds since the causal cones coincide with the null cones of g_{ab} . For $V(\phi) = m^2 \phi^2$ and, more generally, for any non-negative potential the DEC holds with the upshot that (NSP2) and (NSP3) both hold along side (NSP1).

However, negative potentials have been contemplated in cosmological models with dark energy, and for negative potentials DEC and WEC as well can fail. For instance, with $V(\phi) = -m^2 \phi^2$ we get the Klein-Gordon+ equation

$$\nabla^a \nabla_a \phi + m^2 \phi = 0 \tag{KG+}$$

Even though DEC and WEC now fail, giving violations of (NSP2) and (NSP3), (NSP1) holds since there is still a good initial value formulation with the causal cones coinciding with the null cones of g_{ab} .

These examples illustrate that the three criteria of no superluminal propagation can render different verdicts. My claim is that (NSP1) is the criterion to be trusted; in particular, it is sufficient to guarantee no superluminal propagation notwithstanding failures of (NSP2) and (NSP3). Before I defend this position I will address what might seem to be a puzzle about how such a clash of the three criteria could arise.

¹⁵There are no constraints in this case.

6. How can (NSP1) hold when DEC fails?

One might have the intuition that, given the interpretation of the DEC as meaning that the energy flow of matter is never faster than light, the failure of DEC would “mess up” the initial value problem by wrecking uniqueness of solutions. The concern can be put formally without intuition mongering. In the case of the Klein-Gordon equation, uniqueness of solutions can be proved by using the DEC and the conservation condition $\nabla_a T^{ab}(\phi) = 0$ (see Wald 1984, pp. 247-248). Thus, one might worry that for other scalar fields where the DEC fails because, say, the field has a negative potential, the uniqueness proof will break down. The response is that while this particular proof of uniqueness does break down, there are other proofs that use the DEC in a more subtle way. It turns out to be sufficient for uniqueness to associate with the field a symmetric tensor field $S^{ab}(\phi)$ which satisfies DEC. $S^{ab}(\phi)$ need not be identical with the stress-energy tensor of the field, nor need it satisfy the conservation condition $\nabla_a S^{ab}(\phi) = 0$ as long as the failure to be conserved is bounded (see Hawking and Ellis 1972, Section 7.4).

7. Adjudicating the clash of criteria of causal propagation

First a word about semantics. If the energy flow of matter as measured by an observer with four-velocity V^b is defined to be ${}_V P^a := -T^{ab}V_b$ then—by definition—DEC expresses the non-superluminal nature of the energy flow of matter. But the substantive issue here is whether or not DEC is a necessary condition for the non-superluminal propagation of matter-energy. I claim not and, thus, that (NSP2) and (NSP3) fail as explications of no superluminal propagation.

To help make the case consider again a linear scalar field propagating on Minkowski spacetime. It was noted above that for the Klein-Gordon field both DEC and the conservation condition $\nabla_a T^{ab}(\phi) = 0$ are satisfied and, thus, the Hawking and Ellis lemma (4.3.1) implies that (NSP3) is satisfied. But for this case the satisfaction of (NSP3) can be seen more directly, at least when $m > 0$. Evaluating the time-time component T^{44} in an inertial coordinate system gives $\frac{1}{2}[\|\vec{\nabla}\phi\|^2 + \left(\frac{\partial\phi}{\partial t}\right)^2 + m^2\phi^2]$. So when $m > 0$ the vanishing of the stress-energy tensor on a $S(t = \text{const})$ hypersurface implies that ϕ vanishes on $S(t = \text{const})$ as well. But for a linear field (where the superposition of two solutions is a solution), uniqueness requires that if ϕ vanishes

on a region $A \subset S(t = \text{const})$ then it must vanish on $D(A)$; and, a fortiori, $T^{ab}(\phi)$ must vanish on $D(A)$. By contrast, for a $m > 0$ scalar field with negative potential energy, e.g. the Klein-Gordon+ field with $V(\phi) = -m^2\phi^2$, DEC fails and the vanishing of $T^{ab}(\phi)$ on the initial value hypersurface does not entail the vanishing of ϕ (the 44-component of the stress-energy tensor is now $\frac{1}{2}[\vec{\nabla}\phi]^2 + \left(\frac{\partial\phi}{\partial t}\right)^2 - m^2\phi^2$). Thus, although $T^{ab}(\phi)$ vanishes on $A \subset S(t = \text{const})$ the non-null field on A can evolve to produce non-vanishing $T^{ab}(\phi)$ at points of $D(A)$. It should be clear, however, that the violations of (NSP2) and (NSP3) that occur in this case do not herald the superluminal propagation of *anything*. Linearity and uniqueness for the Klein-Gordon+ field imply that whenever the field vanishes on $A \subset S(t = \text{const})$ it must also vanish on $D(A)$. This ensures that the field itself does not propagate superluminally and, a fortiori, that the mass-energy, stress-energy, or anything else carried by the field does not propagate superluminally.

For non-linear fields the vanishing of the field on $A \subset S$ need not entail that it must also vanish on $D(A)$. But the local character of uniqueness part of (NSP1) ensures that the field itself does not propagate superluminally and, a fortiori that the mass-energy, stress-energy, or anything else carried by the field does not propagate superluminally. The argument is simply that the uniqueness part of (NSP1) ensures that with the initial data held fixed on $A \subset S$, the solution on $D(A)$ stays the same however the initial data is jiggled on S/A .¹⁶ (As noted above, an alternative way of making the point is that (NSP1) ensures that, with initial data held fixed on S/A , changing the initial data on A results in solutions that show changes only in the domain of causal influence $J(A)$ of A .) It is, of course, possible to maintain that when DEC fails matter-energy can flow superluminally and, in particular, can flow into $D(A)$ from portions of S/A . And it is further possible to maintain that the fact that no experiment can be performed to detect the effect of this (alleged) superluminal flow is explained by a conspiracy that is enforced by the equations governing the field since they guarantee (NSP1)! As with conspiracy theories in general it is impossible to refute this superluminal flow conspiracy theory. But I submit that here is one place where a mild dose of operationalism has a salutary effect. For those who refuse to swallow

¹⁶If constraints are present then the wiggling of the initial data on S/A while holding fixed the data on A cannot be not arbitrary. This does not affect the KG and KG+ cases since no constraints are present in these cases.

this medicine this much can be said: If (NSP1) holds then, as far as science can tell, the world presents itself as if there is no superluminal propagation despite the fact that (NSP2) and (NSP3) may fail. This “as if” is as good as it gets in science.

To apply these conclusions to a concrete example, consider again a perfect fluid. As measured by an observer comoving with the fluid, the energy flow in the sense of DEC is defined to be ${}_W P^a = \mu W^a$ where W^a is the four-velocity of the fluid/observer. Thus, since W^a is always timelike, ${}_W P^a$ either vanishes (when $\mu = 0$) or else is always timelike regardless of whether the DEC holds. But for a non-comoving observer ${}_V P^a = -(\mu + p)W^a(W_b V^b) - pV^a$, and when DEC fails this four-vector can be spacelike depending on the choice of V^a . Whether the field, and any energy-momentum it carries, propagates non-superluminally is *prima facie* an intrinsic, observer independent matter. (NSP1) makes it so. But, to its detriment, the standard reading of DEC makes it an observer dependent question if DEC is deemed as necessary condition for non-superluminal mass-energy propagation. If the equation of state is such that (NSP1) holds then the present point of view entails that the spacelike character of ${}_V P^a$ does not represent a faster-than-light propagation of matter-energy—in the philosophers’ language, the spacelike curves that thread ${}_V P^a$ represent a pseudo-causal process. Of course, this is just jargon and its use does help to provide a positive account of what ${}_V P^a$ represents. The task of providing such an account will have to await another occasion.

In sum, I maintain that (NSP1) is *sufficient* for no superluminal propagation. If this is correct then neither (NSP2) nor (NSP3) can be *necessary* for no superluminal propagation since (NSP1) can hold when (NSP2) and (NSP3) fail due to a failure of DEC. This point is of no great importance for classical relativistic fields since in most known physical applications in this setting WEC and DEC are in fact satisfied, although there may be some exceptions (see Vollick 1997). However, the point does become important for quantum fields which are known to violate the classical energy conditions at least on small scales (see the following Section).

Finally, note that by present lights the DEC, even in conjunction with the conservation law $\nabla_a T^{ab} = 0$, is not sufficient to guarantee no superluminal propagation. (DEC) and the conservation law are satisfied for a perfect fluid with $\mu \geq |p|$. With an equation of state chosen so that $dF/d\mu > 1$ there is an initial value formulation with the causal cones taken to the sound cones; but since the velocity of sound is greater than the velocity of light the field propagates faster than light.

8. Causal propagation for quantum fields

At the most fundamental level the world is not accurately described by classical relativistic fields but by relativistic quantum fields. Does this mean that the above discussion is of mere academic interest? I claim not; indeed, I claim that the considerations raised the preceding sections take on special importance in QFT.

For some quantum fields a renormalized stress-energy tensor with finite expectation values can be constructed, the quantized Klein-Gordon field being one example (see Wald 1994 for details). However, there is no hope that analogs of even the most basic of the energy conditions imposed on classical relativistic fields—WEC in particular—can be met by their quantized counterparts since it is known that energy density of quantum fields can take on unboundedly negative values at points of spacetime (see Epstein et al. 1965), and this is so even when the counterpart classical fields satisfy both WEC and DEC. Attempts have been made to put bounds on the extent of the violation of energy conditions for quantum fields since large scale violations would lead to macroscopic violations of the laws of thermodynamics, the possibility of creating wormholes that would allow travel to the past, and a parade of other horrors (see, for example, Fewster and Everson 1998 for a derivation of some bounds). However, the success in finding such bounds is only cold comfort for someone who wants to demonstrate that quantum fields do not propagate superluminally even at very small scales but is worried that such a demonstration is made impossible by the violation of energy conditions at these scales.

My prescription for such worries is two-fold. First, absorb the lesson that WEC and DEC are *not necessary* for no superluminal propagation of classical fields. (NSP1) *is sufficient*, and (NSP1) can hold even when WEC and DEC fail. Second, in the context of the algebraic approach to relativistic quantum field theory, recognize that (NSP1) has a natural analog for quantum fields.¹⁷ In the algebraic approach it is postulated that a net of local von Neumann algebras $\mathfrak{R}(\mathcal{O})$ is associated with open bounded regions $\mathcal{O} \subset \mathcal{M}$ of the spacetime manifold \mathcal{M} , where $\mathfrak{R}(\mathcal{O}_1) \subseteq \mathfrak{R}(\mathcal{O}_2)$ if $\mathcal{O}_1 \subseteq \mathcal{O}_2$.¹⁸ Here attention is restricted to the case of QFT on Minkowski spacetime, although some of the

¹⁷See Haag (1992) for an overview of algebraic QFT.

¹⁸A von Neumann algebra \mathfrak{R} is an algebra of bounded operators acting on a Hilbert space \mathcal{H} , its defining feature being that it is closed in the weak topology of \mathcal{H} or, equivalently, $\mathfrak{R} = (\mathfrak{R}')'$ where the prime denotes the commutant.

key points carry over to QFT on globally hyperbolic curved spacetimes. The global algebra $\mathfrak{R}(\mathcal{M}) := \vee_{\mathcal{O}} \mathfrak{R}(\mathcal{O})$ is the algebra generated by the local $\mathfrak{R}(\mathcal{O})$ as \mathcal{O} ranges over all $\mathcal{O} \subset \mathcal{M}$. A state ω on $\mathfrak{R}(\mathcal{M})$ is an expectation value assignment (a normed positive linear functional). Let \mathfrak{S} denote the set of states on $\mathfrak{R}(\mathcal{M})$ that are deemed to be physically possible.¹⁹ Then the quantum analog (QNSP1) of the local form of determinism embodied in (NSP1) is the condition demanding that for any open spacetime region $\mathcal{O} \subset \mathcal{M}$ and any $\omega_1, \omega_2 \in \mathfrak{S}$, if $\omega_1|_{\mathfrak{R}(\mathcal{O})} = \omega_2|_{\mathfrak{R}(\mathcal{O})}$ then $\omega_1|_{\mathfrak{R}(D(\mathcal{O}))} = \omega_2|_{\mathfrak{R}(D(\mathcal{O}))}$, where $\omega|_{\mathfrak{R}}$ denotes the restriction of the state ω to \mathfrak{R} and $D(\mathcal{O})$ is the domain of dependence of \mathcal{O} defined in terms of the the null cones of Minkowski spacetime.²⁰ The choice of the set of physically possible states does not matter much for models that satisfy the AQFT axiom of local primitive causality (first introduced by Haag and Schorer 1962), requiring that for any $\mathcal{O} \subset \mathcal{M}$, $\mathfrak{R}(\mathcal{O}) = \mathfrak{R}(D(\mathcal{O}))$.²¹ For obviously local primitive causality entails entails (QNSP1) for any \mathfrak{S} . It has been verified that local primitive causality holds for the quantized Klein-Gordon field and the Dirac field (see Dimock 1980, 1982).

The upshot is that if (QNSP1) does adequately capture no superluminal propagation for quantum fields, local primitive causality is sufficient to ensure no superluminal propagation even at very small scales despite the failure of various energy conditions at these scales. Admittedly, much more would need to be said in support of affirming the antecedent of this conditional, but I reserve that task for another occasion. Here I simply remark that if the task can be successfully completed then the ideas developed in the preceding Sections have applications that extend far beyond the case of classical relativistic fields.²²

Finally, it is worth remarking that the classical relativistic fields that have been successfully quantized are those that enforce no superluminal propagation. Is this a happenstance? Or does quantum physics somehow serve as a

¹⁹ \mathfrak{S} will, of course, vary from case to case. For the Klein-Gordon field \mathfrak{S} has been identified with the Hadamard states that give a finite expectation value for the renormalized stress-energy tensor (see Wald 1994).

²⁰To get an analog of the global form of classical determinism take \mathcal{O} to be a finite “time sandwich” about a Cauchy surface of Minkowski spacetime.

²¹That $\mathfrak{R}(\mathcal{O}) \subseteq \mathfrak{R}(D(\mathcal{O}))$ follows from the net property of the local algebras. The axiom in question requires the converse.

²²For further aspects of causality in algebraic quantum field theory, see Earman and Valente (2014).

selection principle in favor of no superluminal propagation?

9. Causal vs. pseudo-causal processes

There is a considerable philosophical literature devoted to an analysis of the notion of a “causal process.” The topic goes back at least as far as Bertrand Russell’s *Human Knowledge* (1948), where the concept of a “causal line” is introduced and discussed in somewhat vague terms.²³ Subsequent attempts to add precision include Salmon’s mark transmission theory, conserved quantity theories, transference theories, trope persistence theories, and more (for a good overview and references to the literature, see Dowe 2007). The checkered history of this topic does not show any convergence to a simple and unified account that is adequate to the many different types of examples that might be included under the rubric of causal processes.

The considerations in the preceding Sections suggest a way to handle one of the concerns that drives the causal process literature; namely, how to react to Hans Reichenbach’s examples of processes where *something* moves faster than light (see Reichenbach 1956, Section 23). The issue is illustrated by Wesley Salmon’s spotlight example (Salmon 1984, pp. 141-144). Imagine a spotlight placed at the center of a circular stadium. If the walls are sufficiently distant from the spotlight then even a slowly rotating spotlight will, if sufficiently powerful, cast a spot on the walls that moves faster than light. The almost uniform reaction in the philosophical literature goes as follows: first, it is explicitly or tacitly assumed that relativity theory makes superluminal processes *verboten*²⁴; it is concluded that the spotlight example and its ilk must be defused; and, further, the way to defuse them is dismiss the movement of the spot of light on the walls and its ilk as “unreal sequences” (Reichenbach’s phrase) or a “pseudo-causal processes” (the terminology preferred in the later literature); and finally, it is explicitly or tacitly assumed, to make this resolution work requires a firm grip on what counts as a real sequence or a genuine causal process. Unfortunately, the extant literature on causal processes has not yet provided the grip, so yet more papers are needed.

²³Russell, whose main concern at the time was inductive inference, gave “causal line” an epistemic role whereas later writers envision an ontological role.

²⁴Sometimes it is said that what relativity theory forbids is superluminal causal signals. The notion of signal opens another can of worms since it involves the notion of information transmission. See Weinstein (2006).

No! Unless I have mischaracterized it, this reaction to the Reichenbach examples is both wrong and misguided. It is wrong because the starting presumption that relativistic theories cannot accommodate superluminal processes is false (Geroch 2011). And it is misguided because when dealing with theories that, provably, do not permit superluminal propagation no positive account of “causal process” is needed to dismiss Reichenbach’s examples. This is fortunate because there is good reason for pessimism about the prospects for an approach that seeks to identify genuine causal processes and then defines pseudo-causal processes by negation. The pessimism stems partly from an induction on the failures in causal process literature and partly from the conviction that it is naive to believe that the evolution of fields must be sufficiently tidy as to present us with patterns that are identifiable as genuine causal lines or processes.²⁵

To keep the discussion as simple as possible, assume that the physical systems in question can be described by classical relativistic fields; quantum fields can be handled following the suggestions of Section 8. When the field laws take the form of symmetric hyperbolic pdes that admit an initial value formulation (IVF*) there is an obvious way to pursue the high-road approach I advocate. Two cases need to be considered.

Case 1. The causal cones defined by the system of hyperbolic pdes coincide with the null cones of the spacetime metric (the Minkowski metric η_{ab} for special relativistic theories or the metric tensor g_{ab} of a solution to Einstein’s field equations for general relativistic theories). Then (IVF*) is equivalent to (IVF), and the satisfaction of (NSP1) ensures that field and, hence, the matter-energy it carries does not propagate faster than light—or so I have argued. We have seen that (NSP1) can hold even though the energy-momentum density ${}_V P^a := -T^{ab}V_b$ of the fields as measured by an observer with four-velocity V^b fails to satisfy DEC and, thus, the integral curves of ${}_V P^a$ can have spacelike segments. Such segments can, with justification, be dismissed as pseudo-causal lines, at least insofar as causal lines are supposed to track propagation of the field. More generally, when (NSP1) is satisfied, the motion of a something (e.g. the motion of Salmon’s spotlight on the stadium wall) can be dismissed as a pseudo-causal process if the motion is faster than light. Such negative implications do not by themselves lead to a positive account of what a genuine casual process is. But such an account is

²⁵One could try to argue that we could not long survive unless the portion of the universe we inhabit exhibits patterns readily discernible by us as causal processes.

not needed in the present context to underwrite the labeling of the processes in question as pseudo-causal processes.

Case 2. (IVF*) is not equivalent to (IVF) because the causal cones defined by the system of hyperbolic pdes are “wider” than the null cones of the spacetime metric. Then the presumption of the causal process literature that a relativistic theory cannot accommodate superluminal propagation is proved false. The case of a perfect fluid with an equation of state $1 < dF/d\mu$ is a concrete example (recall Section 5).²⁶ Two subcases need to be considered. (a) The motion of a something whose four-velocity vector lies outside the causal cones. Since (IVF*) holds, the motion of this something can be dismissed as a pseudo-causal process. For just as the satisfaction of (IVF) provides a sufficient condition for no propagation outside the null cones, so the satisfaction of (IVF*) provides a sufficient condition for no propagation outside the causal cones. Again no positive account of genuine causal processes is needed to back the dismissal. (b) The motion of a something whose four-velocity vector lies outside the null cones but not outside the causal cones. Now the high road approach is mute. Additional considerations would have to be marshalled in order to dismiss the motion of such a something in this category as a pseudo-causal process.

In addition to the lacunae it leaves in Case 2(b), my proposed high road approach to the Reichenbach examples can be challenged on various grounds. First, it is limited to cases where the fundamental fields admit an initial value formulation of the form (IVF*). I can only respond that I do not see how it is possible to discuss causal propagation for fields when (IVF*) fails. Second, when (IVF*) holds but (IVF) fails (Case 2 above) and superluminal propagation is possible, there is the worry that a paradox could be generated by rigging the system supporting superluminal propagation to produce inconsis-

²⁶Admittedly, there is some vagueness in what is to count as a “relativistic theory.” To illustrate, one might try to say that a special (respectively, general) relativistic field theory is one which can be properly situated in Minkowski spacetime (respectively, a general relativistic spacetime); and what this means is that the field equations of the theory can be written as pdes using no other spacetime structure than what is definable in terms of the Minkowski metric (respectively, the metric of a general relativistic spacetime). This formulation is intended to rule out smuggling in additional spacetime structure, such as a distinguished frame of reference. Now the issue is what counts as additional spacetime structure vs. another physical field. Fortunately, it not necessary to tackle these issues here, for it suffices to have clear examples of relativistic theories in the intended sense that allow faster-than-light propagation. I take the theory of the perfect fluid in Section 5 to be a clear example. See Geroch (2011) for other examples.

tent closed causal loops. Here a longer response is required, the beginnings of which I will try to supply.

In part, the worry about Case 2 is addressed by considering how the superluminally propagating field couples to other fields. As emphasized by Geroch (2011), the key question here is just a generalized version of the original threshold question; namely, can the coupled field equations be written in the appropriate hyperbolic form so that (IVF*) holds for the total system? If so, then no “rigging” that can be modeled in terms of the interaction of the fields obeying the coupled field equations will result in a paradox since it is guaranteed that consistent initial data will result in a unique consistent evolution of the total system. If it is insisted that, in order to completely assuage the worry about causal paradoxes, it is also necessary to show that no inconsistent loops result from “rigging” by *deus ex machina* interventions on the system supporting superluminal propagation, including interventions that contravene the field equations, then physics per se can be of no help. The trouble with this higher standard is two-fold. First, it is not clear that even theories that do not involve superluminal propagation can yield consistent results under law-contravening *deus ex machina* interventions. Second, the application of such a standard is at the mercy of notoriously interest-relative and context-dependent judgements about what would happen if such-and-such were to be done to the system. I suggest that one should be satisfied with showing that superluminal propagation can be robustly physically consistent in the sense that the system that supports such propagation can interact in a variety of ways with other systems to produce consistent evolutions of the total system for all initial conditions compatible with the combined field equations of the total system.

There are obviously a number of promissory notes in my high road response. But the labor needed to pay them off is of a refreshingly different character than that involved in continuing the game of example-and-counterexample that takes place in the philosophical literature on genuine causal vs. pseudo-causal processes.

10. Conclusion

For classical relativistic fields I have defended a condition—(NSP1)—as sufficient to ensure no superluminal propagation of the fields, and I have indicated that this condition has a natural analog for quantum fields, at least

in the algebraic approach to QFT. It follows that alternative criteria for no superluminal propagation that have been proposed in the literature cannot serve as necessary conditions since these conditions can fail even when (NSP1) holds. These points are well known—to the experts in mathematical physics and a few philosophers of physics—but they seem not to have penetrated general philosophical consciousness and certainly not the literature on causal processes. That literature can (perhaps a bit harshly) be characterized as trying to draw the distinction between genuine vs. pseudo-causal processes using “truisms” about relativity theory, bits of pop physics, and ever more clever examples and counterexamples. I have suggested that part of the goal—finding a basis for classifying the Reichenbach-Salmon examples as examples of pseudo-causal processes—can be achieved in a cleaner and more principled way, at least for systems describable by classical relativistic or relativistic quantum fields, by proving results about the class of solutions of the equations governing the fields. The results reveal that some of the “truisms” about relativity theory are not true.

Acknowledgments

My biggest debt is to Robert Geroch from whom I have learned so much. I am also grateful to Erik Curiel, Giovante Valente, Robert Wald, and Jim Weatherall for helpful conversations and to an anonymous referee whose comments led to important clarifications.

A version of this article will appear in *Studies in History and Philosophy of Modern Physics*.

References

- Dimock, J. (1980). Algebras of local observables on a manifold. *Communications in Mathematical Physics*, 77, 219-228.
- Dimock, J. (1982). Dirac quantum fields on a manifold. *Transactions of the American mathematical Society*, 269, 133-147.
- Dowe, P. (2007). Causal processes. *The Stanford Encyclopedia of Philosophy* <http://plato.stanford.edu/entries/causation-process/>
- Earman, J. and Valente, G. (2014). Relativistic causality in algebraic quantum field theory. *International Studies in Philosophy of Science*, 28, 1-48.

- Epstein, H., Glaser, V., and Jaffe, A. (1965). Nonpositivity of the energy density in quantized field theories. *Nuovo Cimento*, 36, 1016-1022.
- Fewster, C. J. and Everson, S. P. (1998). Bounds on negative energy densities in flat spacetime. *Physical Review D*, 58, 084010-1-6.
- Geroch, R. (1996). Partial differential equations in physics. In G. S. Hall and J. R. Pulham (Eds.), *Proceedings of the Forty-Sixth Scottish Summer School in Physics*. Edinburgh: SUSSP Publications. arXiv: gr-qc/9602055.
- Geroch, R. (2011). “Faster Than Light?” arXiv:1005.1614v1.
- Haag, R. (1992). *Local quantum physics*. Berlin: Springer-Verlag.
- Haag, R. and Schroer, B. (1962). Postulates of quantum field theory. *Journal of Mathematical Physics*, 3, 248-256.
- Hawking, S. F. and Ellis, G. F. R. (1972). *The large scale structure of space-time*. Cambridge: Cambridge University Press.
- Manchak, J. B. (2013). Global spacetime structure. In B. Batterman (Ed.), *The Oxford handbook of the philosophy of physics* (pp. 587-606). Oxford: Oxford University Press.
- Randall, A. D. (2008). *Partial differential equations in general relativity*. Oxford: Oxford University Press.
- Reichenbach, H. (1956). *The philosophy of space and time*. Berkeley: University of California Press.
- Russell, B. (1948). *Human knowledge: Its scope and limits*. New York: Simon and Schuster.
- Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.
- Vollick, D. (1997). How to produce exotic matter using classical fields. *Physical Review D*, 56, 4720–4723.
- Wald, R. M. (1984). *General relativity*. Chicago, IL: University of Chicago Press.
- Wald, R. M. (1994). *Quantum field theory in curved spacetime and black hole thermodynamics*. Chicago, IL: University of Chicago Press.
- Weinstein, S. (2006). Superluminal signaling and relativity. *Synthese*, 148, 381-399.