J.S. Bell's Concept of Local Causality

Travis Norsen
Marlboro College
Marlboro, VT 05344
norsen@marlboro.edu
(Dated: January 5, 2010)

J.S. Bell is universally hailed as a brilliant interpreter and thoughtful critic of quantum theory, and his seminal discovery, Bell's theorem, has been described as "the most profound discovery of science." Yet – although the theorem is now more than 4 decades old and although everyone regards it as profoundly illuminating and although the mathematics involved is quite trivial – there is still rampant controversy about its meaning and implications. Indeed, the vast majority of physicists (including many experts in the foundations of quantum theory) disagree with Bell's own interpretation of his theorem (and are often unaware of this fact). The goal of the present paper is to get behind the controversy by focusing on the theorem's central premise: Bell's concept of relativistic local causality. We collect and organize most of Bell's crucial statements on this topic, which are scattered throughout his writings, into a self-contained, pedagogical discussion, focusing in particular on the concepts of "beables," "completeness," and "causality" and their role in Bell's formulation. An emerging theme of the discussion is the crucial importance of understanding Bell's concept as applying primarily to *candidate theories*.

I. INTRODUCTION

The concept of local causality has played a role in the construction and assessment of physical theories throughout the history of physics. [1, 2] For example, there were important debates between Isaac Newton and some critics of his theory of gravitation which centered on the theory's alleged positing of non-local action-at-a-distance. Newton's own view, interestingly, was that although his theory did claim that (for example) the sun exerted causal influences on the distant planets, this was in principle perfectly consistent with local causality, which he strongly endorsed. His idea was that the theory simply didn't yet provide a complete description of the detailed (and presumably local) mechanism "by and through which [massive bodies'] action and force may be conveyed from one to another." [3]

This response, however, was somewhat empty since at the time nothing was definitely, unambiguously excluded by the requirement of locality. Any apparent action-at-a-distance in a theory could be rendered compatible with local causality by following Newton and simply denying that the theory in question provided a *complete* description of the relevant phenomena.

This changed in 1905 with Albert Einstein's discovery of Special Relativity (SR), which for the first time identified a certain class of causal influences as definitely inconsistent with local causality – namely, those which propagate super-luminally, faster-than-light. As Einstein explained,

> "The success of the Faraday-Maxwell interpretation of electromagnetic action at a distance resulted in physicists becoming convinced that there are no such things as instantaneous action at a distance (not involving an intermediary medium) of the type of Newton's law of gravitation. According to the theory of relativity, action at a distance with the velocity of light always takes the place of instantaneous action at a distance or of action at a distance with an infinite velocity of transmission. This is connected with the fact that the velocity c plays a fundamental role in this theory." [4, pg 47]

The speed of light c plays a fundamental role *vis-a-vis* causality in SR because of the relativity of simultaneity. For two events A and B with space-like separation (i.e., such that a signal connecting A and B would have to propagate super-luminally), the time ordering is ambiguous: different inertial observers will disagree about whether A precedes B in time, or vice versa. There is thus, according to SR, no objective matter of fact about which event occurs first – and hence no possibility of a causal relation between them, since the relation between a cause and its effect is necessarily time-asymmetric. As J.S. Bell put this point, "To avoid causal chains going backward in time in some frames of reference, we require them to go slower than light in any frame of reference." [5, pg 236]

It didn't take long for the relativistic concept of local causality to be used in a criticism of other developing theories. Indeed, it was Einstein himself – in both the famous, but widely misunderstood, EPR paper [6] and several related but less widely known arguments [7] – who first pointed out that the developing Copenhagen quantum theory violated SR's locality constraint. In particular, according to Einstein, that theory's account of *measurement* combined with Niels Bohr's *completeness doctrine* committed the theory to precisely the sort of non-local causation that was, at least according to Einstein, prohibited by SR. Einstein thus rejected Bohr's completeness doctrine and supported (something like) what is now unfortunately [8] called the "local hidden variables" program.

Note the interesting parallel to Newtonian gravity here, with the non-locality in some candidate theory being rendered either real or merely apparent, depending on whether or not one interprets the theory as providing a *complete* description of the physical processes in question. Einstein's assessment of Copenhagen quantum theory with respect to local causality is thus logically parallel to Newton's analysis of his own theory of gravitation.

This brings us to the main subject of the present paper: the work of J.S. Bell. Bell (unlike the vast majority of his, and still our, contemporaries) accepted Einstein's proof of the non-locality of Copenhagen quantum theory. In particular, Bell accepted as valid the EPR-type argument "*from locality to* deterministic hidden variables." [5, pg 157]

The setup for this argument involves a pair of specially-prepared particles which are allowed to separate to remote locations. An observation of some property of one particle then permits the observer to learn something about a corresponding property of the distant particle. According to the Copenhagen view, the distant particle fails to possess a definite value for the property in question prior to observation, and so it is precisely the observation of the nearby particle which (through wave function collapse) brings this newly real property for the distant particle into existence.

But for Einstein, Podolsky, and Rosen – in Bell's summary – this

> "simply showed that [Bohr, Heisenberg, and Jordan] had been hasty in dismissing the reality of the microscopic world. In particular, Jordan had been wrong in supposing that nothing was real or fixed in that world before observation. For after observing only one particle the result of subsequently observing the other (possibly at a very remote place) is immediately predictable. Could it be that the first observation somehow fixes what was unfixed, or makes real what was unreal, not only for the near particle but also for the remote one? For EPR that would be an unthinkable 'spooky action at a distance'. To avoid such action at a distance [one has] to attribute, to the space-time regions in question, *real* properties in advance of observation, correlated properties, which *predetermine* the outcomes of these particular observations. Since these real properties, fixed in advance of observation, are not contained in quantum formalism, that formalism ... is *incomplete*. It may be correct, as far as it goes, but the usual quantum formalism cannot be the whole story." [5, 143]

Thus, Bell agreed with Einstein that the local hidden variables program constituted the only hope for a locally causal re-formulation of quantum theory.

Bell's legendary contribution, however, was a theorem establishing that no such local hidden variable theory – *and hence no local theory of any kind* – could generate the correct empirical predictions for a certain class of experiment. [5] According to Bell, we must therefore accept the real existence, in nature, of faster-than-light causation – in apparent conflict with the requirements of SR:

> "For me then this is the real problem with quantum theory: the apparently essential conflict between any sharp formulation and fundamental relativity. That is to say, we have an apparent incompatibility, at the deepest level, between the two fundamental pillars of contemporary theory..." [5, pg 172]

Indeed, Bell went so far as to suggest, in response to his theorem and the related experimental data, the rejection of "fundamental relativity" and the return to a Lorentzian view in which there is a dynamically privileged (though probably empirically undetectable) reference frame:

> "It may well be that a relativistic version of [quantum] theory, while Lorentz invariant and local at the observational level, may be necessarily non-local and with a preferred frame (or aether) at the fundamental level." [5, pg 33]

And elsewhere:

> "...I would say that the cheapest resolution is something like going back to relativity as it was before Einstein, when people like Lorentz and Poincare thought that there was an aether – a preferred frame of reference – but that our measuring instruments were distorted by motion in such a way that we could not detect motion through the aether. Now, in that way you can imagine that there is a preferred frame of reference, and in this preferred frame of reference things do go faster than light. ....Behind the apparent Lorentz invariance of the phenomena, there is a deeper level which is not Lorentz invariant... [This] pre-Einstein position of Lorentz and Poincare, Larmor and Fitzgerald, was perfectly coherent, and is not inconsistent with relativity theory. The idea that there is an aether, and these Fitzgerald contractions and Larmor dilations occur, and that as a result the instruments do not detect motion through the aether – that is a perfectly coherent point of view." [9]

Our intention is not to lobby for this view here, though it is unfortunate that most physicists today think (erroneously) that such a view has been empirically refuted. We mention it only to stress how seriously Bell took his

own interpretation of his own theorem. He was so convinced of the need for super-luminal causation that he was willing to seriously contemplate and even advocate something most physicists would consider unthinkable: that we've all, for about 100 years now, fundamentally misunderstood the meaning or status of relativity theory.

Since Bell's untimely death in 1990, the experimental data verifying the relevant predictions of quantum mechanics has grown quite strong. [10] Thus, the "essential conflict" Bell spoke of is no longer merely between relativistic local causality and the *predictions* of quantum theory; the conflict is between Bell's local causality and *experiment*. Thus, to whatever extent one accepts Bell's local causality concept as an appropriate statement of the requirements of relativity theory, one must evidently join Bell in seriously re-assessing the status and foundations of relativity.

There is some wiggle room here in the sense that there now exist toy models which, despite containing the required violations of Bell's local causality, are nevertheless plausibly "relativistic." [15] Bell initiated the research in this direction and, based on his preliminary findings, backed away somewhat in his final years from the claim that a blatant violation of the principles of relativity, such as a preferred frame or "aether," is required by the experiments.

Nevertheless, the need to worry about the reconciliation of quantum theory and relativity is much stronger than most physicists appreciate. This is because most commentators on Bell and Bell's theorem interpret the theorem (plus the associated experiments) not as refuting a plausibly relativistic concept of local causality, but rather as refuting something else. The claim, typically, is that Bell's concept of local causality smuggles something in that doesn't belong, and hence fails either to appropriately capture the causal structure required by relativity, or to respect certain principles allegedly mandated by quantum theory. Hence the widespread claims that Bell's work has refuted not locality but the *hidden variables* program, or *determinism*, or micro-physical *realism*.[11–14]

In some ways, it is puzzling that so many radically different interpretations of the meaning of Bell's theorem could exist – especially considering that the mathematics involved in the derivation of an empirically testable Bell-type inequality from Bell's local causality concept is completely straightforward and simple. But it isn't the derivation itself which is controversial. What's controversial is rather the exact nature and meaning of the premise – Bell's concept of local causality – from which that derivation proceeds.

The goal of the present paper, therefore, is to explicate and illuminate this crucial concept. Being sympathetic to Bell's case, we think the best way to clarify and defend it is to let Bell speak for himself. The present paper thus collects and organizes Bell's various statements on this topic (which are scattered about his various papers) with the hope of injecting Bell's actual views back into
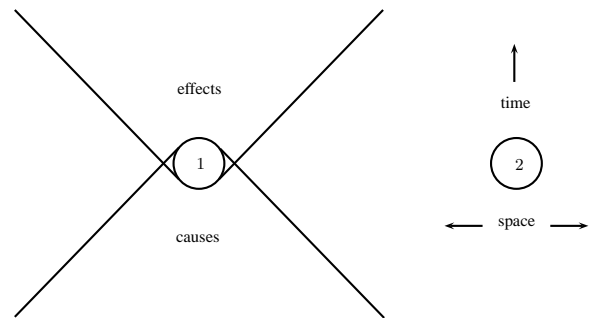

FIG. 1: "Space-time location of causes and effects of events in region 1." (Figure and caption are from [5, pg 239].)

the ongoing discussion of these issues. In particular we hope that this paper will serve as a more appropriate first introduction to these issues (than the usual textbook accounts, which contain errors and misrepresent Bell's views) for students.

It should be understood at the outset that virtually all of the points discussed here are included because some kind of misunderstanding of them has been present and influential in the Bell literature. We will provide citations to works which we think exemplify the various important misunderstandings. But length considerations (and the desire to keep this a self-contained, positive presentation of Bell's views) forbid any extensive polemical discussions. The unfortunate implication is that it may not always be obvious to the reader how (or even exactly where) a given author makes the kind of mistake we are alleging. This should not, however, be seen as a shortcoming of the present essay, whose purpose, after all, is precisely to re-ignite discussion of these issues with a more careful eye on Bell's own views.

The paper is organized as follows. In the following section, we jump quickly from some of Bell's preliminary, qualitative statements to his final, quantitative formulation of relativistic local causality. Then, in a series of topically-organized subsequent sections, we will highlight and explore various aspects by clarifying some perhaps-unfamiliar or suspicious concepts which appear in Bell's formulation and by contrasting them to various other ideas with which they have sometimes been confused.

II. LOCAL CAUSALITY: OVERVIEW

Let us begin with a qualitative formulation of Bell's concept of local causality. In a 1988 interview, in answer to the question "What does *locality* mean?" Bell responded:

> "It's the idea that what you do has consequences only nearby, and that any consequences at a distant place will be weaker and will arrive there only after the time permitted by the velocity of light. Locality is

the idea that consequences propagate continuously, that they don't leap over distances." [17]

Bell gave a slightly more careful (but still qualitative) formulation of what he calls the "Principle of local causality" in a 1990 paper ("La nouvelle cuisine"):

> "The direct causes (and effects) of events are near by, and even the indirect causes (and effects) are no further away than permitted by the velocity of light." [5, pg 239]

Then, citing a figure which we have reproduced here as Figure 1, Bell continues:

> "Thus for events in a space-time region 1 ... we would look for causes in the backward light cone, and for effects in the future light cone. In a region like 2, space-like separated from 1, we would seek neither causes nor effects of events in 1." [5, pg 239]

This should be uncontroversial as a statement of the constraints (on the causal relations between events) which follow from the light-cone structure attributed to space-time in SR. Bell immediately notes, however, that "[t]he above principle of local causality is not yet sufficiently sharp and clean for mathematics."

Here, then, is Bell's sharpened and cleaned formulation of special relativistic local causality. (The reader should remember that this is, at this point, merely a 'teaser' which those not already familiar with it should only expect to understand after further reading.)

> "A theory will be said to be locally causal if the probabilities attached to values of local beables in a space-time region 1 are unaltered by specification of values of local beables in a space-like separated region 2, when what happens in the backward light cone of 1 is already sufficiently specified, for example by a full specification of local beables in a space-time region 3..." [5, page 239-40]

The space-time regions referred to are illustrated in Figure 2. We may translate Bell's formulation into mathematical form as follows:

$$P(b_1|B_3, b_2) = P(b_1|B_3), \qquad (1)$$

where b_i refers to the value of some particular beable in space-time region i and B_i refers to a *sufficient* (for example, a complete) specification of all beables in the relevant region. The Ps here are the probabilities assigned to event b_1 by the candidate theory in question. Equation 1 thus asserts mathematically just what Bell states in the caption of his accompanying figure (reproduced here as Figure 2): "full specification of [beables] in 3 makes events in 2 irrelevant for predictions about 1 in a locally causal theory."

Let us then jump right in to a closer examination of the several puzzling features of this formulation.
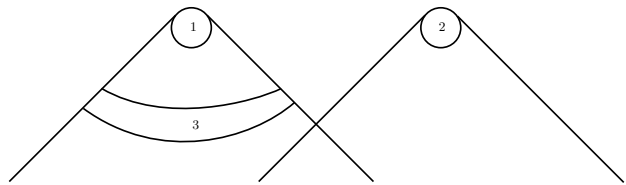

FIG. 2: "Full specification of what happens in 3 makes events in 2 irrelevant for predictions about 1 in a locally causal theory." (Figure and caption are from [5, pg. 240].)

III. BEABLES

For those to whom the term is new, the first question about the word "beable" is: how to pronounce it? The word has three syllables. It does not rhyme with "feeble," but with "agreeable." Bell invented the word as a contrast to the "observables" which play a fundamental role in the formulation of orthodox quantum theory, so let us begin there.

A. Beables vs. Observables

Beables (as contrasted to observables) are those elements of a theory which (according to the theory) correspond to something that is *physically real*, independent of any observation. Bell elaborates:

> "The beables of the theory are those elements which might correspond to elements of reality, to things which exist. Their existence does not depend on 'observation'. Indeed observation and observers must be made out of beables." [5, pg 174]

Or as he explains elsewhere,

> "The concept of 'observable' .... is a rather woolly concept. It is not easy to identify precisely which physical processes are to be given the status of 'observations' and which are to be relegated to the limbo between one observation and another. So it could be hoped that some increase in precision might be possible by concentration on the *be*ables ... because they are there." [5, pg 52]

Bell's reservations here (about the concept "observable" appearing in the fundamental formulation of allegedly fundamental theories) are closely related to the so-called "measurement problem" of orthodox quantum mechanics, which Bell famously encapsulated by remarking that the orthodox theory is "unprofessionally vague and ambiguous" [5, pg 173] in so far as its fundamental dynamics is expressed in terms of "words which, however legitimate and necessary in application, have no place in a *formulation* with any pretension to physical precision" –

such words as "*system, apparatus, environment, microscopic, macroscopic, reversible, irreversible, observable, information, measurement.*" [5, pg 215] As Bell elaborates,

> "The concepts 'system', 'apparatus', 'environment', immediately imply an artificial division of the world, and an intention to neglect, or take only schematic account of, the interaction across the split. The notions of 'microscopic' and 'macroscopic' defy precise definition. So also do the notions of 'reversible' and 'irreversible'. Einstein said that it is theory which decides what is 'observable'. I think he was right – 'observable' is a complicated and theory-laden business. Then the notion should not appear in the *formulation* of fundamental theory. *Information? Whose* information? Information about *what?*" [5, pg 215]

According to Bell, the orthodox versions of quantum theory are simply unclear ("vague and ambiguous") on the question of what is really there, what is actually, *physically* real, independent of observation. Bell thus believed that one needed a better theory which was not ambiguous in this way.

It is crucial to appreciate that the goal here is not to insist on some particular type of entity (particles, fields, hidden variables, whatever) being physically real in the face of orthodox quantum theory's claims to the contrary. (And there is no question about the correctness of quantum theory's empirical *predictions.*) Rather, the point is simply to insist on clarity about the ontological status of elements in the theory, *as posited by the theory*:

> "The terminology, *be*-able as against *observable*, is not designed to frighten with metaphysic those dedicated to realphysic. It is chosen rather to help in making explicit some notions already implicit in, and basic to, ordinary quantum theory. For, in the words of Bohr, 'it is decisive to recognize that, however far the phenomena transcend the scope of classical physical explanation, the account of all evidence must be expressed in classical terms.' It is the ambition of the theory of local beables to bring these 'classical terms' into the equations, and not relegate them entirely to the surrounding talk." [5, pg 52]

Bell is here pointing out that even Bohr (a convenient personification of skepticism regarding the physical reality of unobserved microscopic phenomena) recognizes certain things (for example, the directly perceivable states of a classical measuring apparatus) as unambiguously physically real, i.e., as beables. The unprofessional vagueness and ambiguity of orthodox quantum theory, then, is related to the fact that its formulation presupposes these beables, but fails to provide clear mathematical laws to describe them. As Bell explains,

> "The kinematics of the world, in [the] orthodox picture, is given by a wavefunction ... for the quantum part, and classical variables – variables which *have* values – for the classical part... [with the classical variables being] somehow macroscopic. This is not spelled out very explicitly. The dynamics is not very precisely formulated either. It includes a Schrödinger equation for the quantum part, and some sort of classical mechanics for the classical part, and 'collapse' recipes for their interaction." [5, pg 228]

So there are two related problems. First, the posited ontology is rather different on the two sides of (what Bell calls) "the shifty split" [5, pg 216] – that is, the division between "the quantum part" and "the classical part." But then, as a whole, the posited ontology remains unavoidably vague so long as the split remains shifty. And second, the *interaction* across the split is problematic. Not only is the account of this dynamics (the "collapse" process) inherently bound up in concepts from Bell's list of dubious terms, but the very existence of a special dynamics for the interaction seems to imply conflicts – inconsistencies – with the dynamics already posited for the two realms separately.

As Bell summarizes elsewhere,

> "I think there are *professional* problems [with quantum mechanics]. That is to say, I'm a professional theoretical physicist and I would like to make a clean theory. And when I look at quantum mechanics I see that it's a dirty theory. The formulations of quantum mechanics that you find in the books involve dividing the world into an observer and an observed, and you are not told where that division comes... So you have a theory which is fundamentally ambiguous..." [9]

The point of all this is to clarify the sort of theory Bell had in mind as satisfying the relevant standards of professionalism in physics. It is often thought, by those who do not understand or do not accept Bell's criticisms of orthodox quantum theory, that the very concept of "beable" (in terms of which his concept of local causality is formulated) commits one already to hidden variables or determinism or realism or some other physically or philosophically dubious principle.

But this is not correct. The requirement here, ultimately, is only that candidate fundamental theories – at least, those "with any pretension to physical precision" [5, pg 215] – be formulated clearly and precisely. And this requires, according to Bell, that the theories provide a *uniform* and *consistent* candidate description of physical reality. In particular, there should be no ambiguity regarding what a given candidate theory is fundamentally *about* (the beables), nor regarding precisely how those posited physically real elements are posited to act and interact (the laws).

B. Beables vs. Conventions

So far we have explained the term "beable" by contrasting it to the "observables" of orthodox quantum theory. We must now also contrast the concept of "beables" to those elements of a theory which are, to some degree, conventional:

> "The word 'beable' will also be used here to carry another distinction, that familiar already in classical theory between 'physical' and 'non-physical' quantities. In Maxwell's electromagnetic theory, for example, the fields **E** and **H** are 'physical' (beables, we will say) but the potentials **A** and ϕ are 'non-physical'. Because of gauge invariance the same physical situation can be described by very different potentials. It does not matter [i.e., it is not a violation of local causality] that in Coulomb gauge the scalar potential propagates with infinite velocity. It is not really supposed to *be* there. It is just a mathematical convenience." [5, pg 52-3]

Or, as Bell puts the same point in another paper,

> "...there *are* things which *do* go faster than light. British sovereignty is the classical example. When the Queen dies in London (long may it be delayed) the Prince of Wales, lecturing on modern architecture in Australia, becomes *instantaneously* King.... And there are things like that in physics. In Maxwell's theory, the electric and magnetic fields in free space satisfy the wave equation
>
> $$\frac{1}{c^2}\frac{\partial^2 \mathbf{E}}{\partial t^2} - \nabla^2 \mathbf{E} \;=\; 0$$
> $$\frac{1}{c^2}\frac{\partial^2 \mathbf{B}}{\partial t^2} - \nabla^2 \mathbf{B} \;=\; 0$$
>
> ...corresponding to propagation with velocity c. But the scalar potential, if one chooses to work in 'Coulomb gauge', satisfies Laplace's equation
>
> $$-\,\nabla^2 \phi = 0$$
>
> ...corresponding to propagation with infinite velocity. Because the potentials are only mathematical conveniences, and arbitrary to a high degree, made definite only by the imposition of one convention or another, this infinitely fast propagation of the Coulomb-gauge scalar potential disturbs no one. Conventions can propagate as fast as may be convenient. But then we must distinguish in our theory between what is convention and what is not." [5, pg 234]

Thus, in order to cleanly decide whether a given theory is or is not consistent with local causality,

> "you must identify in your theory 'local *be*ables'. The *be*ables of the theory are those entities in it which are, at least tentatively, to be taken seriously, as corresponding to something real. The concept of 'reality' is now an embarrassing one for many physicists.... But if you are unable to give some special status to things like electric and magnetic fields (in classical electromagnetism), as compared with the vector and scalar potentials, and British sovereignty, then we cannot begin a serious discussion." [5, pg 234]

This explains why, according to Bell: "It is in terms of local beables that we can hope to formulate some notion of local causality." [5, pg 53]

C. Beables and Candidate Theories

It is important to appreciate that a beable is only a beable relative to some particular candidate theory which posits those elements as physically real (and, presumably, gives precise mathematical laws for their dynamics). For example, the fields **E** and **B** (and not the potentials) are beables according to classical Maxwellian electrodynamics as it is normally understood. But one could imagine some alternative theory which (in response, say, to something like the Aharanov-Bohm effect) posits the potentials as beables (and hence also presumably posits some "one true gauge").

Thus, we must separate any questions about what the "real beables" are into two distinct parts: first, what elements does a given candidate theory posit as beables; and second, which candidate theory do we think is *true*? The point is, you don't have to be able to answer the second question in order to answer (for a given theory) the first. This should provide some comfort to those who (perhaps influenced by positivist or instrumentalist philosophy) think it out of the question that we could ever discover that a given theory (at least, of the sort Bell considers professional) is true; such people could still accept Bell's characterization of when such "a theory will be said to be locally causal."

But even those who are not skeptical on principle recognize that, because of the complexity in practice of settling questions about the truth status of scientific theories, some tentativeness is often in order. Bell recognizes this too:

> "I use the term 'beable' rather than some more committed term like 'being' or 'beer' to recall the essentially tentative nature of any physical theory. Such a theory is at best a *candidate* for the description of nature. Terms like 'being', 'beer', 'existent',

etc., would seem to me lacking in humility. In fact 'beable' is short for 'maybe-able'." [5, pg 174]

The crucial point is that the "maybe" here pertains to the epistemological status of a given candidate theory. By contrast, the "beable status" of certain elements of a theory *relative to that theory* should be completely straightforward and uncontroversial. If there is any question about what elements a theory posits as beables, it can only be because the proponents of the theory have not (yet) sufficiently clarified what the theory is about, what the theory *is*. Whether the theory is true or false is an orthogonal question.

All of that said, Bell does take certain elements largely for granted as beables – that is, as beables that *any* serious candidate theory would have to recognize as such: "The beables must include the settings of switches and knobs on experimental equipment, the currents in coils, and the readings of instruments." [5, pg 52] As noted before, even someone like Bohr must evidently concede the real existence (the beable status) of these sorts of things. And, more philosophically, since our primary cognitive access to the world is through "switches and knobs on experimental equipment" and other such directly perceivable facts – i.e., since such facts must always constitute the primary *evidence* on which we will have to rest any argument for the truth of a particular candidate physical theory – it is hard to imagine a serious theory which doesn't grant such facts beable status. A theory which didn't would evidently have to regard our perception as systematically delusional, and hence would have to regard any alleged empirical evidence – for anything, including itself – as delusional. In short, such a theory would necessarily be self-refuting. [18]

We stress this point for two related reasons. First, anyone who is uncomfortable with the apparently "metaphysical" positing of ultimate "elements of reality" (even in a tentative way, through the tentative positing of a candidate physical theory) should be relieved to find that the concept "beable" is merely a placeholder for whatever entities we (tentatively) include in the class which *already*, by necessity, exists and includes certain basic, directly-perceivable features of the world around us. And second, these particular beables – e.g., the settings of knobs and the positions of pointers – have a particularly central role to play in the derivation (from Bell's concept of local causality) of the empirically testable Bell inequalities.

D. What about non-local beables?

Let us finally note the word "local" which appears in Bell's formulation of local causality in the phrase "local beables." Bell explains: "*Local* beables are those which are definitely associated with particular space-time regions. The electric and magnetic fields of classical electromagnetism ... are again examples." [5, pg 234-5] Sim-

ilarly, in another paper, Bell stresses that the "local beables of [a] theory [are] associated with definite positions in space." [5, pg 176] The contrast here is evidently beables that are not local, such as the wave function of ordinary quantum mechanics (a function not on 3-space but on a much higher-dimensional configuration space).

Some questions are raised by the fact that Bell's definition of local causality mentions only the local beables of the theory in question. Is the implication that theories positing non-local beables are to be dismissed as already violating the constraints of relativistic causality, on that basis alone? Then Bell's careful definition of local causality would be needed (to assess the locality of the laws posited by the theory) only when a theory passed the primary test of positing *exclusively local* beables. [19, 20]

It is by no means clear that Bell had this in mind, though it does seem like a plausible reading of Bell's formulation. After all, non-local beables (by definition) do not even live on the spacetime "stage" posited by SR. Assuming a theory's non-local beables play some causal role in the theory (and if not, there would be no point positing them) it then follows that a theory positing non-local beables posits causal influences between a given physical event in space-time and "something" that is literally outside the special-relativistic physical universe. This hardly seems compatible with the causal structure, illustrated in Figure 1, we began by assuming SR required.

On the other hand, a completely literal reading of Bell's local causality condition would have it permitting a theory to posit non-local beables and still, in principle, come out as a locally causal theory. The problem is, theories that posit non-local beables do so for a reason – namely, those beables play a crucial role in the theories' dynamics for the other, local beables out of which things like experimental apparatus pointers are evidently built. So it is very difficult to know quite what is meant when one asks such a theory to define a probability for a certain happening (b_1) in region 1, on the basis of – exclusively – the *local* beables in region 3. Perhaps such a probability could be generated by averaging over all possible configurations for the non-local beables that are (according to the theory) consistent with whatever is specified about the local beables. But this seems poorly-defined and (worse) against the basic spirit of what Bell has suggested, which is that a full or complete specification of local beables in region 3 should (in locally causal theories) specify the causes of events in region 1 sufficiently enough that events in 2 are rendered redundant or irrelevant. In a theory positing non-local beables, B_3 (which remember we are for the moment taking to refer only to the local beables of the theory) will necessarily fail to sufficiently capture the causes of events in 1.

For our purposes, however, this interpretational debate can be safely bracketed. For there is a sense in which the quantum mechanical wave function is "local enough" – enough, that is, to treat it as a local beable for the purposes of applying Bell's local causality criterion.

Because Bell's condition involves a complete specifica-

tion of (local) beables across (what might as well be) a space-like hypersurface, the wave function "fits" in this space-time region and so we may still include it as part of the "complete specification of local beables" in region 3. Then there are no ambiguities: even including the wave function in this way, theories like Orthodox QM and Bohmian Mechanics *do manifestly violate* Bell's locality condition. Indeed, Bell himself takes exactly this approach: "it is notable that in this argument nothing is said about the locality, or even localizability, of the variable λ [which is, in the context of this statement, a subset of what we have here called B_3]. These variables could well include, for example, quantum mechanical state vectors, which have no particular localization in ordinary space-time." [5, pg 153]

A final point specifies the context in which the idea of locality (including both the distinction between locally causal and non-local theories, and that between local and non-local beables) is meaningful: as Bell notes,

> "it may well be that there just *are* no local beables in the most serious theories. When space-time itself is 'quantized', as is generally held to be necessary, the concept of locality becomes very obscure. And so it does also in presently fashionable 'string theories' of 'everything'. So all our considerations are restricted to that level of approximation to serious theories in which space-time can be regarded as given, and localization becomes meaningful." [5, pg 235]

IV. COMPLETENESS

Having clarified the concept of "beables" which appears in Bell's formulation of local causality, let us now turn to the last phrase in that formulation, italicized here:

> "A theory will be said to be locally causal if the probabilities attached to values of local beables in a space-time region 1 are unaltered by specification of values of local beables in a space-like separated region 2, *when what happens in the backward light cone of 1 is already sufficiently specified, for example by a full specification of local beables in a space-time region 3...*" [5, page 239-40]

In a word, the key assumption here is "that events in 3 be specified *completely*" [5, pg 240] (emphasis added).

Let us first see why this requirement is necessary. Consider again Figure 2, and suppose that $\bar{B}_3$ denotes an *incomplete* specification of beables in region 3. It can then be seen that a violation of

$$P(b_1|\bar{B}_3, b_2) = P(b_1|\bar{B}_3) \qquad (2)$$

does *not* entail the existence of any super-luminal causal influences. For suppose some event "X" in the overlapping backwards light cones of regions 1 and 2 causally influences both b_1 and b_2. It might then be possible to infer, from b_2, something about X, from which one could in turn infer something about b_1. Suppose, though, that the incomplete description of events in region 3 – $\bar{B}_3$ – omits precisely the "traces" of this past common cause X. Then b_2 could usefully supplement $\bar{B}_3$ – i.e., Equation 2 could be violated, in the presence of purely local causation.

Thus, as Bell explains, in order for Equation 1 to function as a locality criterion,

> "it is important that events in 3 be specified completely. Otherwise the traces in region 2 of causes of events in 1 could well supplement whatever else was being used for calculating probabilities about 1. The hypothesis is that any such information about 2 becomes redundant when 3 is specified completely." [5, pg 240]

And here is the same point from an earlier paper:

> "Now my intuitive notion of local causality is that events in 2 should not be 'causes' of events in 1, and vice versa. But this does not mean that the two sets of events should be uncorrelated, for they could have common causes in the overlap of their backward light cones [in a local theory]. It is perfectly intelligible then that if [B_3] in [region 3] does not contain a complete record of events in that [region], it can be usefully supplemented by information from region 2. So in general it is expected that [$P(b_1|b_2, \bar{B}_3) \neq P(b_1|\bar{B}_3)$.] However, in the particular case that [B_3] contains already a *complete* specification of beables in [region 3], supplementary information from region 2 could reasonably be expected to be redundant." [5, pg 54]

It is important to stress that, like the concept of beables itself, the idea of a sufficient (full or complete) specification of beables is relative to a given candidate theory. What Bell's local causality condition requires is that – *in order to assess the consistency between a given candidate theory and the relativistic causal structure sketched in Figure 1* – we must include, in B_3, everything *that candidate theory* says is present (or relevant) in region 3. It is not, by contrast, necessary that we achieve omniscience regarding what *actually exists* in some spacetime region.

The appearance of the word "completeness" tends to remind commentators of the EPR argument, and hence apparently also tends to suggest that Bell smuggled into his definition of local causality the unwarranted assumption that orthodox quantum theory is incomplete. (See,

for example, Ref. [23].) As mentioned earlier, Bell did accept the validity of the EPR argument. But this means only that, according to Bell, local causality (plus some of QM's predictions) entail the incompleteness of orthodox QM. His view on that point, however, is no part of his *formulation* of local causality. This should be clear, for example, from the fact that the formulation nowhere mentions any particular candidate beables, such as the wave function (which is supposed in orthodox QM to provide, all by itself, a complete specification of beables, at least on the microscopic side of the shifty split). Nevertheless, this vague linguistic association (between Bell's requirement that "events in 3 be specified completely" and the conclusion of the EPR argument) has been a major driving force behind the decades-long misconception that Bell's work refutes not local causality, but only the hidden variables program.

What Bell's locality concept does say is only this: *whatever* your theory posits as physically real (in region 3), make sure you include all of that when calculating the relevant probabilities to test whether your theory respects or violates Equation 1, i.e., whether your theory is or isn't locally causal in the sense of Figure 1. No assumptions are made about the type of theory to which the locality criterion can be applied. In particular, the incompleteness of orthodox quantum theory (i.e., the existence of "hidden variables") is not assumed. The virtue of Bell's formulation lies precisely in this generality. [24]

V. CAUSALITY

Recall the transition from Bell's preliminary, qualitative formulation of local causality to the final, "sharp and clean" version. And recall in particular Bell's statement that the preliminary version was *insufficiently* sharp and clean for mathematics. What is it, exactly, that Bell considered inadequate about the qualitative statement? It seems likely that it was the presence there of the terms "cause" and "effect" which are notoriously difficult to define mathematically. Indeed, about his final formulation Bell says: "Note, by the way, that our definition of locally causal theories, although motivated by talk of 'cause' and 'effect', does not in the end explicitly involve these rather vague notions." [5, pg 240]

How exactly does Bell's "definition of locally causal theories" fail to "explicitly involve" the "rather vague notions" of cause and effect? On its face this sounds paradoxical. But the resolution is simple: what Bell's definition actually avoids is any specific commitment about what physically exists and how it acts. As mentioned above, any such commitments would restrict the generality of the locality criterion, and hence undermine the profound scope of Bell's theorem. Instead, Bell's definition shifts the burden of providing some definite account of causal processes *to theories* and itself merely defines a space-time constraint that must be met if the causal processes *posited by a candidate theory* are to be deemed

local in the sense of SR.

The important mediating role of candidate theories *vis-a-vis* causality will be further stressed and clarified in the following subsection. Subsequent subsections further clarify the concept of "causality" in Bell's "local causality" by contrasting it with several other ideas with which it has often been confused or conflated.

A. Causality and candidate theories

As already discussed, according to Bell it is the job of physical theories to posit certain physically real structures (beables) and laws governing their interactions and evolution. Thus Bell's definition of locally causal theories is not a specification of locality for a particular type of theory, namely, those that are "causal" – with the implication that there would exist also theories that are "non-causal." A theory, by the very nature of what we mean by that term in this context, is automatically causal. "Causal theory" is a redundancy. And so, as noted earlier, one must understand Bell's "definition of locally causal theories" as a criterion that theories – i.e., candidate descriptions of causal processes in nature – must satisfy in order to be in accord with special relativistic locality. In short, the causality in the "definition of locally causal theories" is simply whatever a given candidate theory says, about whatever it says it about.

As Bell explains, the practical *reason* for defining local causality in terms of the physical processes posited by some candidate theory (as opposed to the physical processes which actually exist in nature) has to do with our relatively direct access to the one as opposed to the other:

> "I would insist here on the distinction between analyzing various physical theories, on the one hand, and philosophising about the unique real world on the other hand. In this matter of causality it is a great inconvenience that the real world is given to us once only. We cannot know what would have happened if something had been different. We cannot repeat an experiment changing just one variable; the hands of the clock will have moved, and the moons of Jupiter. Physical theories are more amenable in this respect. We can *calculate* the consequences of changing free elements in a theory, be they only initial conditions, and so can explore the causal structure of the theory. I insist that [the theory of local beables, i.e., the local causality concept] is primarily an analysis of certain kinds of physical theory."[5, pg 101]

It is worth stressing that, as was noted earlier for "beables," the fact that the "causality" in Bell's concept of local causality is not specified (but is rather left for individual candidate theories to define for themselves) re-

moves the ground from under those who worry that Bell is smuggling in some extra, perhaps "metaphysical," requirement here. In particular, even those philosophical skeptics who think it is impossible that we could ever empirically determine which candidate theory is true, need not worry. As Bell points out, we needn't believe that a given theory is really true in order to analyze its posited causal structure (and in particular the consistency of that posited structure with SR).

Who could possibly disagree with any of this? Since philosophers revel in their inability to define causality, one sometimes sees claims that violations of Bell's locality concept (or the resulting inequality) indicate only the existence of some non-local correlations, with a final judgment about whether any non-local *causation* occurs having to wait on some further philosophical explication of "causation" (which often involves Reichenbach's Principle of the Common Cause). [26, 27]

This is starkly at odds with Bell's own view, which was that no such philosophical account of causation is needed, since physical theories – *qua* claims about what exists (beables) and how those beables act and interact (laws) – are *ipso facto* making causal claims. That is, Bell seemed to think (and I agree) that it is a trivial matter to decide, for some (unambiguously formulated) candidate physical theory, what is and is not a genuinely *causal* influence. That is, we can directly "explore the causal structure of" the candidate theory. And so the suggestion that (in the face of empirical evidence violating the inequality which follows from Bell's concept) we might retain locality and reject the idea of causality, is thus nonsensical. It amounts to an attempt to bypass the central role Bell attributed to *candidate theories* in formulating the very idea of local causality.

Having argued that there is no such thing as a theory that is not "causal" (and hence that the "causality" in "local causality" is simply whatever beables and processes are posited by the theory whose locality we are assessing with Bell's locality criterion) we should point out also that there is no real meaning to "local" other than "locally causal." In the context of relativity theory, "local" just *means* that physical goings-on in a space-time region are (causally) influenced only by physical goings-on in the past light cone, and (causally) influence physical goings-on only in the future light cone, as sketched in Figure 1.

The general point here is that we should not understand the word "causal" as adding anything to an allegedly more minimal "local theory." There is no such thing as a theory that isn't a "causal theory" and no such thing as a theory that is "local" but not "locally causal." Bell's "definition of locally causal theories" is, in the final analysis, simply a criterion for deciding whether the physical processes posited by a given candidate *theory* are *local*, in the sense of respecting the relativistic structure shown in Figure 1.

B. Causality vs determinism

The previous subsection stressed that the "causal" in "locally causal theories" simply refers to the physically real existents and processes (beables and associated laws) posited by some candidate theory, whatever exactly those might be. We in no way restrict the class of theories (whose locality can be assessed by Bell's criterion) by introducing "causality." In particular, the word "causal" in "locally causal theories" is not meant to imply or require that theories be *deterministic* as opposed to irreducibly stochastic.

> "We would like to form some [notion] of *local causality* in theories which are not deterministic, in which the correlations prescribed by the theory, for the beables, are weaker." [5, pg 53]

Bell thus uses the word "causal" quite deliberately as a wider abstraction which *subsumes but does not necessarily entail* determinism.

This is manifested most clearly in the fact that Bell's mathematical formulation of "local causality" – Equation 1 – is stated in terms of probabilities. Indeed, in "The Theory of Local Beables" [5, pg 52-62] Bell discusses "Local Determinism" *first*, arguing that, in a "local deterministic" theory the actual *values* of beables in region 1 (of Figure 2) are determined by a complete specification of beables in region 3 (with additional specification of beables from region 2 being redundant). In our mathematical notation, local determinism means

$$b_1(B_3, b_2) = b_1(B_3) \tag{3}$$

(or perhaps it is sufficient that the right hand side be defined at all!) where, as before, b_1 and b_2 are the values of specific beables in regions 1 and 2, while B_3 denotes a sufficient (e.g., complete) specification of beables in region 3.

In a (local) stochastic theory, however, even a complete specification of causally relevant beables (those in region 3 of Figure 2) may not determine the realized value of the beable in question (in region 1). Rather, the theory specifies only *probabilities* for the various *possible* values that might be realized for that beable. Of course, determinism is a special case of the more general assumption of stochasticity.

> "Consider for example Maxwell's equations, in the source-free case for simplicity. The fields **E** and **B** in region 1 are completely determined by the fields in region 3, regardless of those in 2. Thus this is a locally causal theory in the present sense. The deterministic case is a limit of the probabilistic case, the probabilities becoming delta functions." [5, pg 240]

The natural generalization of the above mathematical formulation of "local determinism" is precisely Bell's local causality condition:

$$P(b_1|B_3, b_2) = P(b_1|B_3), \qquad (4)$$

i.e., b_2 is irrelevant – not for determining *what happens* in region 1 because that, in a stochastic theory, is simply not determined – but rather for determining the *probability* for possible happenings in region 1. Such probabilities are the "output" of stochastic theories in the same sense that the actual realized values of beables are the "output" of deterministic theories. Thus, Bell's local causality condition for stochastic theories (Equation 4) and the analogous condition (Equation 3) for deterministic theories, are imposing precisely the same locality requirement on the two kinds of theories: information about region 2 should be redundant (and hence irrelevant) in regard to what the theory says about region 1, once the beables in region 3 are sufficiently specified.

Of course, if one insists that any stochastic theory is *ipso facto* a stand-in for some (perhaps unknown) underlying deterministic theory (with the probabilities in the stochastic theory thus resulting not from indeterminism in nature, but from our ignorance), Bell's locality concept would make no sense. The requirement of a complete specification of beables in region 3 would then contradict the allowance that such a specification does not necessarily determine the events in region 1. But this is no objection to Bell's concept of local causality. Bell is not asking us to accept that any particular theory (stochastic or otherwise) is *true*; he's just asking us to accept his definition of what it would *mean* for a stochastic theory to respect relativity's prohibition on superluminal causation. And this requires us to accept, at least in principle, that there could be such a thing as a genuinely, irreducibly stochastic theory, and that the way "causality" appears in such a theory is that certain beables do, and others do not, influence the probabilities for specific events.

We have been stressing here that "causality" is wider than, and does not necessarily entail, determinism. Bell has deliberately and carefully formulated a local causality criterion that does not tacitly assume determinism, and which is thus stated explicitly in terms of probabilities – *the fundamental, dynamical probabilities assigned by stochastic theories to particular happenings in space-time*. Note in particular that the probabilities in Equation 1 are not subjective (in the sense of denoting the degree of someone's belief in a proposition about b_1), they cannot be understood as reflecting partial ignorance about relevant beables in region 3, and they do not (primarily) represent empirical frequencies for the appearance of certain values of b_1. They are, rather, the fundamental "output" of some candidate (stochastic) physical theory.

C. Causality vs correlation

Everyone knows that correlation doesn't imply causality. Two events (say, the values taken on by beables b_1 and b_2 in Bell's spacetime regions 1 and 2, respectively) may be correlated without there necessarily being any implication that b_1 is the cause of b_2 or vice versa.

> "Of course, mere *correlation* between distant events does not by itself imply action at a distance, but only correlation between the signals reaching the two places." [5, pg 143]

And Bell describes the theme of "La nouvelle cuisine" as "the problem of formulating ... sharply in contemporary physical theory" "these notions, of cause and effect on the one hand, and of correlation on the other". [5, pg 232]

There is a widespread belief that Bell's local causality condition is really only a "no correlation" requirement, such that the empirical violation of the resulting inequalities establishes only "non-local correlations" (as opposed to non-local causation). But this is a misconception. Bell uses the term "causality" (e.g., in talking about his "definition of locally causal theories") to highlight that a violation of this condition (by some theory) means that the theory posits non-local causal influences, *as opposed to* mere "non-local correlations."

It will be clarifying to illustrate this by relaxing several points that Bell has carefully built into his formulation of local causality, and showing that violation of the resulting, weakened conditions may still entail *correlations* between space-like separated events, but no longer implies that there are faster-than-light causal influences. We have done this already, in the previous section, when we explained why a violation of Equation 2 would not (unlike a violation of Equation 1) entail any violation of the causal structure of Figure 1. We now consider a second modified version of Bell's criterion.

Consider again the spacetime diagram sketched in Figure 2. Bell notes that

> "It is important that region 3 completely shields off from 1 the overlap of the backward light cones of 1 and 2." [5, pg 240]

Why is this so important? For example, why couldn't we replace region 3 of Figure 2 with a region like that labelled 3* in Figure 3? This region, just like 3 in Figure 2, closes off the back light cone of 1 and hence – it might seem – would be perfectly sufficient for defining the probabilites associated with b_1 in a local theory.

But a more careful analysis shows that a violation of

$$P(b_1|B_{3^*}, b_2) = P(b_1|B_{3^*}) \qquad (5)$$

(the same as Equation 1 but with region 3 of Figure 2 replaced by region 3* of Figure 3) does *not* entail any non-local causation. Here, there is a perfectly *local* causal mechanism by which b_1 and b_2 can be correlated, in a way
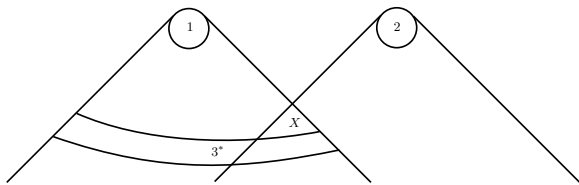
FIG. 3: Similar to Figure 2, except that region 3* (unlike region 3 of Figure 2) fails to shield off region 1 from the overlapping backward light cones of regions 1 and 2. Thus, (following the language of Figure 2's caption) even full specification of what happens in 3* *does not make* events in 2 irrelevant for predictions about 1 in a locally causal theory.

FIG. 4: Full specification of what happens in region 3 *does not make* events in *region 4* irrelevant for predictions about region 1 in a locally causal theory.

that isn't "screened off" by B_{3^*}, thus violating Equation 5 in a situation which involves no violation of special relativistic causation. The mechanism is this: in a stochastic theory, an event may occur at the space-time point labelled "X" in Figure 3 which was *not determined* by the complete specification of beables (B_{3^*}) in region 3*. But despite not having been determined by beables in its past, that event really comes into existence and may in principle have *effects* throughout its future light cone – which includes *both* region 1 *and* region 2. Event X may, so to speak, broadcast sub-luminal influences which bring about correlations between b_1 and b_2, such that information about b_2 is *not* redundant in regard to defining what happens in region 1. Thus we may have a violation of Equation 5 – i.e., the candidate theory in question could attribute different values to $P(b_1|B_{3^*}, b_2)$ and $P(b_1|B_{3^*})$ – despite there being, according to the theory in question, no non-local causation at work. Thus, while Equation 5 may validly be described as a "no correlations" condition for regions 1 and 2, it definitely fails as a "no causality" condition.

If we return, however, to the original region 3 (of Figure 2) which *does* "completely [shield] off from 1 the overlap of the backward light cones of 1 and 2" [5, pg 240] it becomes clear that no such correlation-without-non-local-causality can occur. Here, if some X-like event (not determined by even a complete specification of beables in region 3) occurs somewhere in the future light cone of region 3, it will necessarily fail to lie in the overlapping past light cones of regions 1 and 2 (which would be necessary for it to in turn locally influence both of those events).

Bell has carefully set things up so that a violation of Equation 1 entails that there is some non-local causation. It isn't necessarily that something in region 2 is causally influencing something in region 1, or vice versa. It is always possible that there is some other event, neither in region 1 nor region 2, which was not determined by B_3, and which itself causally influences both b_1 and b_2. The point is, though, that *this* causal influence would have to be non-local (i.e., would have to violate the special relativistic causal structure sketched in Figure 1). [28, pg 130]
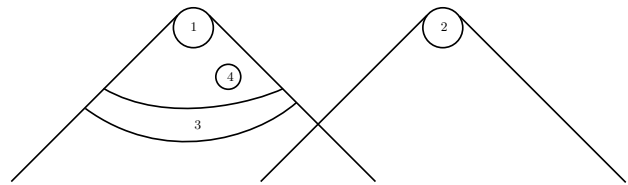
It is important to understand that the locality condition here is not that the probability assigned to b_1 is uniquely or finally or fully or ultimately determined by B_3 alone. Consider, for example, the spacetime region labeled 4 in Figure 4. As with the event X discussed above, in a stochastic theory events may occur in region 4 which are not determined even by a complete specification of beables in region 3. And yet, clearly, information about such events *may* (in a locally causal theory) usefully supplement B_3 in an assignment of probabilities to various possible b_1. In other words, a violation of

$$P(b_1|B_3, b_4) = P(b_1|B_3) \qquad (6)$$

(the same as Equation 1 but with the space-like separated beable b_2 replaced with a beable b_4 from the region 4 sandwiched between 1 and 3) would involve no relativistically forbidden non-local causation. Violation of this condition would show only what we have already assumed – that we are dealing with a stochastic, non-deterministic candidate theory – and would certainly not establish that the theory violated local causality.

Seeing this helps clarify the meaning of Equation 1. The claim made in the local causality condition is not that the probabilities assigned (to events in region 1, on the basis of complete information about region 3) are the "best possible" probabilities the theory allows. They aren't. Better ones might be assigned, e.g., if we move region 3 forward in time, into the more recent past of region 1. The claim is only that (however good they are to begin with) the probabilties assigned to b_1 on the basis of B_3 *do not change upon additional specification of* – i.e., the probabilities *are not affected by* – happenings at spacelike separation.

Earlier in the paper from which that statement is taken, Bell has used the example of a correlation between the ringing of a kitchen alarm and the readiness of a boiling egg. That the alarm rings just as the egg is finished cooking obviously does not entail or even suggest that the ringing *caused* the egg to harden. Correlation does not imply causality. As Bell completes the point,

> "The ringing of the alarm establishes the readiness of the egg. But if it is already given that the egg was nearly boiled a second before, then the ringing of the alarm makes the readiness no more certain." [5, pg 240]

Reading b_2 for "the ringing of the alarm," b_1 for "the readiness of the egg," and B_3 for "the egg was nearly boiled a second before," we have a simple intuitive example of Equation 1: although b_1 and b_2 may be *correlated* such that information about b_2 can tell us something about b_1, that information will be redundant (in a locally causal theory) once B_3 is specified. That is, Equation 1 will be respected. This helps illustrate that what Equation 1 excludes is non-local causation – not correlation. Any theory violating the condition necessarily provides a non-local causal explanation for any predicted correlation between b_1 and b_2.

D. Causality vs signaling

We have stressed that the "causality" in Bell's "local causality" refers to the beables and laws posited by some candidate theory, and have carefully contrasted this "causality" to both determinism and correlation. In this final sub-section, we contrast causality with one final concept with which it is often confused: *signaling.*

"Signaling" refers to a certain human activity, in which one person transmits information, across some distance, to another person. Such transmission clearly requires a causal connection between the sending event and the receiving event. But more is required as well: the ability of the people to send and receive the information. And, in turn, a theoretical description of these acts (sending and receiving) presupposes some account, in a given candidate theory, of what sorts of beables it is or isn't possible for humans to *control.*

In the typical EPR-Bell setup, we have separated observers (traditionally Alice and Bob) making spin-component measurements (using, say, Stern-Gerlach devices oriented spatially along the $\hat{a}$ and $\hat{b}$ directions, respectively) on each of a pair of spin-entangled particles. The outcomes of their individual measurements (manifested in the final location of the particle, or the position of some pointer, or some fact about some other beable) may be denoted A and B respectively.

The beables pertaining to a given run of the experiment may then be cataloged as in Figure 5. Roughly, we may think of $\hat{a}$ and $\hat{b}$ as referring to the (controllable) orientation of the two pieces of measuring apparatus (this being the basis for the notation), and λ as referring to the (not necessarily fully known and not necessarily fully controllable) state of the particle pair emitted by the source. But at this level of abstraction where we are deliberately not committing to any particular theory, that is the best we can do and even that perhaps goes too far (e.g., by talking about "particles").

What is important is just that $\hat{a}$ and λ together contain a sufficient (e.g., complete) specification of beables in a region ($3a$) whose relation to the measurement outcome A is just the relation region 3 has to region 1 in Figure 2. The beables $\hat{b}$ and B are spacelike separated from the measurement outcome A. It is thus a straightforward
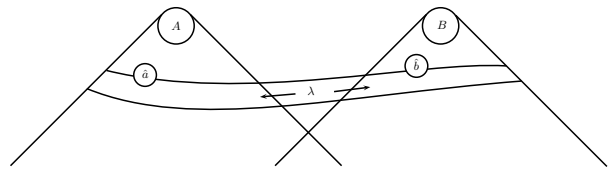

FIG. 5: Space-time diagram illustrating the various beables of relevance for the EPR-Bell setup. Separated observers Alice and Bob make spin-component measurements (using apparatus characterized by variables $\hat{a}$ and $\hat{b}$ respectively) of a pair of entangled spin-1/2 particles. The state of the particles (and/or any other appropriately associated beables) is denoted λ, and the outcomes of the two measurements are represented by the beables A (in region 1) and B (in region 2). Note that λ and $\hat{a}$ collectively constitute a complete specification of beables in space-time region $3a$, which shields off region 1 from the overlap of the past light cones of 1 and 2. Likewise, λ and $\hat{b}$ collectively constitute a complete specification of beables in space-time region $3b$, which shields off region 2 from the overlap of the past light cones of 2 and 1. Thus the joint specification of λ and $\hat{a}$ will – in a locally causal theory – make $\hat{b}$ and B redundant for a theory's predictions about A (and likewise, specification of λ and $\hat{b}$ will render $\hat{a}$ and A redundant for B).

application of Bell's locality criterion that, in a locally causal theory, we should have

$$P(A|\hat{a}, \hat{b}, B, \lambda) = P(A|\hat{a}, \lambda), \qquad (7)$$

the corresponding condition on B

$$P(B|\hat{a}, \hat{b}, A, \lambda) = P(B|\hat{b}, \lambda), \qquad (8)$$

and, consequently, the mathematical condition on the joint probability that is usually called "factorization":

$$P(A, B|\hat{a}, \hat{b}, \lambda) = P(A|\hat{a}, \lambda) \times P(B|\hat{b}, \lambda) \qquad (9)$$

since, e.g., the spacelike-separated facts $\hat{b}$ and B should be dynamically irrelevant to what happens at A, since "what happens in the backward light cone of $[A]$ is already sufficiently specified, for example by a full specification of local beables in" an appropriate region [5, pg 240] (and correspondingly for B). Bell notes:

> "Very often such factorizability is taken as the starting point of the analysis. Here we have preferred to see it not as the *formulation* of 'local causality', but as a consequence thereof." [5, pg 242-3]

What would be required for Bob to be able to send a signal to Alice (using this setup)? We follow Bell and

> "[s]uppose we can control variables like a and b above, but not those like A and B. I do not quite know what 'like' means here, but suppose that beables somehow fall into two classes, 'controllables' and 'uncontrollables'. The latter are no use for *sending* signals, but can be used for *reception*. [5, pg 60]

Then the prohibition of superluminal signaling – signal locality – is the requirement that the *empirical frequency* for Alice's outcome (A) shouldn't depend on the space-like separated beables that Bob can control ($\hat{b}$). This can be stated as a mathematical analog of Bell's locality condition:

$$F(A|\hat{a}, \hat{b}) = F(A|\hat{a}). \tag{10}$$

where F is the predicted empirical frequency for the outcome A, given in terms of the fundamental stochastic-theoretical probabilities by

$$F(A|\hat{a}, \hat{b}) = \sum_B \int P(\lambda) \, P(A, B|\hat{a}, \hat{b}, \lambda) \, d\lambda \tag{11}$$

where $P(\lambda)$ is the probability distribution for states λ associated with whatever preparation procedure is used to produce the states. Bell characterizes this Signal Locality condition as follows:

> "That is to say that, when averaged over the unknown λ, manipulation of b has no effect on the statistics of A..." [5, pg 245]

It is true that a violation of Equation 10 by a theory would mean that the theory supports (i.e., predicts the possibility of) superluminal signaling.

What is crucial here is that "local signaling" is a distinct, and much weaker, condition than "local causality". Equation 7 mathematically entails (using also the definition in Equation 11, Bayes' relation for the joint probability $P(A, B)$, and the fact that $\sum_B P(B|\hat{b}, \lambda) = 1$) Equation 10. That is, any locally causal theory will necessarily also exhibit signal locality, i.e., locally causal theories will prohibit superluminal signalling.

But the converse does not hold: theories exhibiting signal locality (i.e., theories according to which it is not possible to signal superluminally) are not necessarily locally causal! For example, deterministic theories which violate local causality will support superluminal signaling only if λ (whatever that might be for a given theory) is itself sufficiently *controllable*. If, for example, $P(\lambda)$ can be made a delta function, then, in effect, the empirical frequencies are identical to the basic stochastic-theoretical probabilities (there being no longer any non-trivial averaging over the unknown λ). But if λ is not controllable, then the underlying superluminal causation may turn out to be unusable for signalling.

The issue of signalling comes up because there are many who, suspicious of the allegedly metaphysical character of certain aspects of Bell's concept of local causality, reject Bell's criterion and instead insist that the only prohibition special relativity places on candidate theories is that they should forbid non-local (i.e., faster than light) signaling. Bell repudiates this maneuver, however:

> "Do we then have to fall back on 'no signaling faster than light' as the expression of

the fundamental causal structure of contemporary theoretical physics? That is hard for me to accept." [5, pg 245]

Part of his reasoning is as follows:

> "Could ... no-superluminal-signaling ... be regarded as an adequate formulation of the fundamental causal structure of physical theory? I do not think so. For ... the concepts involved in relating it to causal structure are not very satisfactory." [5, pg 238]

For Bell, "signal locality" is of no fundamental interest vis-a-vis the consistency of theories with SR, for reasons that are reminiscent of his dissatisfaction with such notions as "measurement" appearing in the formulation of physical theories. [5, pg 213-231] That is, because it is a uniquely human activity, a clear formulation of "signaling" would seem to require a detailed theory of human activities to be built on top of a given candidate physical theory (specifying, e.g., what limitations exist on their abilities to freely control certain beables posited by the theory):

> "Suppose that we are finally obliged to accept [a theory which is not locally causal]. Can *we* then signal faster than light? To answer this we need at least a schematic theory of what *we* can do, a fragment of a theory of human beings." [5, pg 60]

Or as Bell explains elsewhere,

> "...the 'no signaling...' notion rests on concepts which are desperately vague, or vaguely applicable. The assertion that 'we cannot signal faster than light' immediately provokes the question:
>
> Who do we think *we* are?
>
> *We* who can make 'measurements', *we* who can manipulate 'external fields', *we* who can 'signal' at all, even if not faster than light? Do *we* include chemists, or only physicists, plants, or only animals, pocket calculators, or only mainframe computers?" [5, pg 245]

Implementing the proposed criterion would thus, according to Bell, be a "formidable challenge". [5, pg 155]

A perhaps stronger argument against the proposed criterion is the fact that there are extant theories (which satisfy all the relevant demands of seriousness and professionalism and clarity) which are signal local and yet nevertheless blatantly at odds with relativistic local causality. For example, one may consider the de Broglie - Bohm theory, in which

> "...the consequences of events at one place propagate to other places faster than light.

This happens in a way that we cannot use for signaling. Nevertheless it is a gross violation of relativisitic causality." [5, pg 171]

To Bell, it was obvious that the pilot wave theory violated local causality, despite the fact that, according to that theory, it is impossible for humans to signal faster than light. This illustrates once again that there is a crucial difference between causality and signaling, with the fundamental being, for Bell, "local causality" as against "signal locality."

In terms of understanding Bell's theorem and its implications, however, the important thing is not deciding whether "consistency with relativity" should be gauged in terms of Bell's local causality, or (alternatively) signal locality. For one's position on that issue would in no way affect Bell's proof that no locally causal theory can be in agreement with experiment; it would only render this proof *uninteresting* to the extent that one regarded theories like Bohmian Mechanics and orthodox QM (with their different, but equally blatant, violations of local causality) as nevertheless already perfectly consistent with SR.

The important thing, rather, is the *failure* to carefully distinguish between local causality and signal locality – i.e., the problem is the equivocation between these two concepts. Such equivocation is one of the primary devices by which commentators distort the meaning of Bell's theorem and the implications of the results of the associated experiments. [29]

VI. SUMMARY AND CONCLUSIONS

Many contemporary discussions of Bell's theorem assert outright falsehoods, e.g., that determinism or hidden-variables are explicit premises of the theorem. The better discussions highlight the role of local causality, but still often present a mere formal skeleton of Bell's local causality concept (e.g., by simply beginning with Equation 9) and hence invite erroneous inferences such as that the locality criterion is a mere statistical no-correlations requirement having no substantial relation to the relativistic causal structure of Figure 1, or that it somehow already tacitly presupposes determinism or the incompleteness of orthodox QM.

We have hopefully shown here that there is much more to Bell's concept of local causality. In particular, we have stressed that:

- The locality criterion does not presuppose "hidden variables".

- The locality criterion does not presuppose determinism.

- A theory's violation of the criterion means that it posits non-local *causation*, not mere non-local correlations.

- A theory's violation of the criterion does not necessarily mean that it supports super-luminal signaling

Particularly noteworthy is the plausibility, generality, and evident appropriateness of Bell's locality concept as an expression of the relativistic causal structure of Figure 1. Given that the famous Bell-type inequalities follow as a logical consequence of Bell's concept of local causality, and given that this inequality is violated by the data in actual experiments, it follows that no locally causal theory can correctly describe what is happening in these experiments. And that in turn implies that some non-local, faster-than-light causal influences – in violation of the structure sketched in Figure 1 – actually exist in nature.

Does this mean that relativity is false and that we must adopt a fundamentally un-relativistic approach such as the aether theories mentioned in the introduction? Not necessarily, though almost certainly we should follow Bell and consider that option quite seriously. [30] Should we instead work harder to find a new formulation of quantum theory which, though violating Bell's local causality and the causal structure of Figure 1, manages to do so in a relativistically invariant way, like the model introduced in Ref. [15]? Will we find some unexpected resolution of these tensions in general relativity or quantum theories of gravity?

Our goal here, however, is not to answer any of these questions – only to raise them. Our conclusion, in short, is that such questions are not crazy and are not philosophy. They are, rather, legitimate questions that physicists should, in light of Bell's work, acknowledge as sensible and begin to work to address.

[1] Mary B. Hesse, *Forces and Fields: The Concept of Action at a Distance in the History of Physics*, Dover Publications, Mineola, New York, 2005

[2] Ernan McMullin, "The Explanation of Distant Action: Historical Notes," in James T. Cushing and Ernan McMullin, Editors, *Philosophical Consequences of Quantum Theory*, University of Notre Dame Press, Indiana, 1989

[3] Isaac Newton's famous 25 February 1693 letter to Richard Bentley, quoted (e.g.) in Gerald Holton and Stephen Brush, *Physics: The Human Adventure*

[4] Albert Einstein, *Relativity: The Special and the General Theory*, Penguin Classics, 2006

[5] John S. Bell, *Speakable and Unspeakable in Quantum Mechanics*, 2nd ed., Cambridge University Press, 2004.

[6] Albert Einstein, Boris Podolsky, and Nathan Rosen, "Can quantum-mechanical description of reality be considered complete?" *Phys. Rev.* **47**, 777-780, 1935

[7] Travis Norsen, "Einstein's Boxes", *American Journal of*

Physics, 73(2), 164-176 (2005)

[8] The terminology of "hidden variables" is unfortunate because, at least in the one clear extant example of a "hidden variables theory," the "hidden variables" are precisely the variables which are not, in fact, hidden. About this theory (the de Broglie - Bohm "pilot wave" theory, or "Bohmian Mechanics," which adds to the standard quantum mechanical wave function *definite particle positions* obeying a deterministic evolution law) Bell remarked: "it would be appropriate to refer to the **x**s as 'exposed variables' and to ψ as a 'hidden variable'. It is ironic that the traditional terminology is the reverse of this." [5, pg 128] Or similarly: "Although [in Bohmian Mechanics] Ψ is a real field it does not show up immediately in the result of a single 'measurement,' but only in the statistics of many such results. It is the de Broglie - Bohm variable X that shows up immediately each time. That X rather than Ψ is historically called a 'hidden' variable is a piece of historical silliness." [5, pg 162-3] It is also relevant that the wave function ψ is "hidden" – in the sense of being *not accessible* via experiment – even in orthodox quantum theory (which is suppsed to be the primary example of a non-hidden-variable theory). For some illuminating further discussion, see: Roderich Tumulka, "Understanding Bohmian Mechanics: A Dialogue" *American Journal of Physics*, 72(9), 1220-1226 (2004)

[9] P.C.W. Davies and J.R. Brown, eds., *The Ghost in the Atom*, Chapter 3: interview with J.S. Bell, Cambridge University Press, 1986

[10] Gregor Weihs *et al.*, "Violation of Bell's inequality under strict Einstein locality conditions," *Physical Review Letters*, 81, 5039-5043 (1998)

[11] Eugene P. Wigner, "Interpretation of Quantum Mechanics," 1976, reprinted in *Quantum Theory and Measurement*, J.A. Wheeler and W.H. Zurek, eds., Princeton University Press, 1983.

[12] N. David Mermin, "Hidden Variables and the Two Theorems of John Bell," *Rev. Mod. Phys.*, 65(3), July 1993, pp. 803-815.

[13] Jon Jarrett, "Bell's Theorem: A Guide to the Implications" in J. Cushing and E. McMullin, eds., *Philosophical Consequences of Quantum Theory*, University of Notre Dame Press, Indiana, 1989

[14] Anton Zeilinger, "The message of the quantum," *Nature*, 438, 743 (8 December, 2005)

[15] Roderich Tumulka, "A Relativistic Version of the Ghirardi-Rimini-Weber Model" *J. Stat. Phys.* **125** (2006) 821-840

[16] L.E. Ballentine and Jon P. Jarrett, "Bell's Theorem: Does quantum mechanics contradict relativity?" *American Journal of Physics* **55**(8), 696-701 (1987)

[17] Charles Mann and Robert Crease, "John Bell, Particle Physicist" (Interview), *Omni*, Vol. 10 No. 8, 84-92 and 121 (May 1988)

[18] T. Norsen, "Against 'Realism'," *Foundations of Physics*, **37**(3), 311-340 (2007)

[19] It is certainly possible that a theory which posits only local beables could violate Bell's local causality condition, for the laws governing the evolution and interaction of the beables could be non-local. (Newton's theory of gravity, with a gravitational potential field obeying Poisson's equation, would be an example; see T. Norsen, "The Theory of (Exclusively) Local Beables", arxiv 0909.4553, for such a formulation of quantum theory.) Bell's local

causality condition can thus be understood as a necessary condition for the locality of a theory's *laws* (which condition already takes for granted that the beables posited by the theory are local) – though as noted in the main text it is not entirely clear that Bell interpreted his condition this way. For some illuminating further discussion of the distinction between the locality of beables and laws, see: Tim Maudlin, "Completeness, Supervenience, and Ontology", *J. Phys. A: Math. Theor.* **40** 3151-3171 (2007)

[20] Note that a theory's positing of non-local beables is related to its violating what is sometimes called "separability" or "particularism" in the literature. See, e.g., Paul Teller, "Relational Holism," and Don Howard, "Holism, Separability, and the Metaphysical Implications of the Bell Experiments," in Cushing *et al.* (eds.), *op cit.* (These authors' identification of "separability" with Jarrett's "completeness" condition, however, is incorrect.) John Earman ["Locality, Nonlocality, and Action at a Distance: A Skeptical Review of Some Philosophical Dogmas," in *Kelvin's Baltimore Lectures and Modern Theoretical Physics*, Robert Kargon and Peter Achinstein, *eds.*, MIT Press, Massachusetts, 1987] names a similar condition *pre-locality*, noting that the "locality assumptions built into it seem just as necessary for paradigm cases of physical action at a distance as for local action by contact." This, I think, is correct, and is consistent with the interpretation of Bell's locality condition offered here. But generally speaking, the question of how to understand "locality" (and how to assess the consistency with SR) in theories positing non-local beables is an important and subtle one which deserves further attention.

[21] Once one realizes that the notion of a complete specification of beables is relative to a given candidate theory, there is no further problem understanding the meaning of "complete" or "full". (There is only the problem of deciding which theory is true!) But it is less clear, even given some well-defined candidate theory, what partial specifications of the beables might be considered "sufficient." Bell hints toward an answer here: "In a more careful discussion the notion of completeness should perhaps be replaced by that of sufficient completeness for a certain accuracy, with suitable epsilonics." [5, pg 104]

[22] For example, the vast literature on "counter-factual definiteness" (as it relates to Bell's theorem) has its origins in just this confusion. The critics here evidently think that every mathematical element in the derivation of a Bell-type inequality must have some directly measured (or measurable) empirical counterpart. Yet the derivations inevitably deal simultaneously with pairs such as $A(\hat{a}, \lambda)$ and $A(\hat{a}', \lambda)$ which refer to the outcomes an experiment would have under two mutually exclusive experimental arrangements (denoted by $\hat{a}$ and $\hat{a}'$). And so, the objection goes, we are thus necessarily smuggling in the idea of definite measurement outcomes for non-performed experiments (i.e., "counter-factual definiteness") which is at odds with orthodox quantum philosophy and equivalent to the assumption of determinism and/or hidden variables. What the critics miss, however, is that there has already been an argument "*from locality to*" [5, pg 157] these very deterministic hidden variables. In short, before arriving at the stage of the derivation that the critics criticize, Bell has already proved that viable locally causal theories must have a certain structure: namely, they must contain functions like $A(\hat{a}, \lambda)$

above. And then, as Bell tersely dispatches the earliest objection of this sort: "The quantities $A(\hat{a}', \lambda)$, $B(\hat{b}', \lambda)$ are just the same functions $A(\hat{a}, \lambda)$, $B(\hat{b}, \lambda)$ with different arguments." [From "Reply to Critics"... need to look up the exact reference...] For some further discussion see: T. Norsen, "Bell Locality and the Nonlocal Character of Nature," Foundations of Physics Letters, 19(7), 633-655 (Dec. 2006).

[23] J.F. Clauser, M.A. Horne, A. Shimony, and R.A. Holt, "Proposed Experiment to Test Local Hidden-Variable Theories," *Phys. Rev. Lett.* 23, 880 (1969)

[24] For an alternative formulation of locality which is highly proprietary, i.e., not general like Bell's, see: Abner Shimony, "Bell's Theorem," *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/fall2008/entries/bell-theorem/

[25] It is, however, possible to prove that any locally causal account of certain predictions of quantum theory must include some beables beyond the wave function – i.e., some "hidden variables". This is precisely the EPR argument, which – still – most commentators on Bell's theorem fail to appreciate is a valid argument which plays a crucial logical role in Bell's derivation of the inequality. As Bell reported: "My own first paper on this subject (*Physics* **1**, 195 (1965).) starts with a summary of the EPR argument *from locality to* deterministic hidden variables. But the commentators have almost universally reported that it begins with deterministic hidden variables." [pg 157]

[26] Jeremy Butterfield, "Bell's Theorem: What it Takes," *British Journal for the Philosophy of Science* 42, pp. 41-83

[27] H. R. Brown, *Physical Relativity*, Oxford University Press, 2005

[28] Tim Maudlin, *Quantum Non-Locality and Relativity*, Second Edition, Blackwell (2002)

[29] Greenstein and Zajonc, *The Quantum Challenge*, 2nd edition, Jones and Bartlett, 2005

[30] Tim Maudlin, "Non-local correlations in quantum theory: how the trick might be done" in *Einstein, Relativity and Absolute Simultaneity*, Routledge, 2007