Peaceful Coexistence or Armed Truce?

Quantum Nonlocality and the Spacetime View of the World

by Kent A. Peacock

A Thesis submitted in conformity with the requirements
for the Degree of Doctor of Philosophy in the
University of Toronto.

To the memory

of my father

Donald Knight Peacock

1920 – 1986

ABSTRACT

This thesis is concerned with a critical examination of the notion of “peaceful coexistence” between quantum mechanics and the theory of relativity. This phrase, coined by Abner Shimony, is meant to suggest that quantum mechanics does not predict any observable conflict with the relativistic picture of causation, even though quantum mechanics, a fundamentally nonlocal theory, does clearly violate a condition called “outcome independence” which would seem to be implicit in any local realistic theory.

The main justification that Shimony and others have cited for this view is the No-Signalling Theorem, the claim that quantum nonlocality cannot be exploited to violate the relativistic picture of causation by sending controllable signals outside the light cone. (This would be a violation of what Shimony and Jarrett have called “parameter independence”.) I examine the various proofs that have been given of this theorem, and show that they are essentially circular in the sense that they either incorporate such strong assumptions about the localizability or commutativity of observables as to render them incapable of dealing with the very cases they should be best equipped to treat, or else simply presume some condition equivalent to relativistic causality. Quantum mechanics has therefore not been shown to provide a categorical prohibition against violations of relativity; the question of peaceful coexistence remains open.

I argue that this circumstance should not be viewed as surprising since relativity, being a classical theory, should be expected to be only an approximation. In conclusion I sketch some requisites for a genuinely quantal theory of spacetime.

Acknowledgements

The completion of a doctoral thesis is the culmination of a stage of a person's life, sometimes many years in length, during which time the thesis was not merely researched and written but lived. It would be as difficult to thank everyone who has earned thanks for having helped with my thesis, as it would be to thank everyone who has made my life worth living or even possible over the last five years. But the impossible task cannot be shirked. Here is a list of people—in alphabetical order, to avoid any danger of invidious comparisons—who have in some important way aided this project:

> Eva Broad, Susan and Peter Campbell, J. W. Crichton, P. H. Eberhard, R. A. Imlay, F. L. Jackson, Kitty, John D. Milton, Calvin Normore, Barbara J. Peacock, Jeffrey G. Peacock, Lewis Donald Peacock, Anna M. Pezacki, Thomas M. Robinson, Fay and Al Rooke, William Seager, John G. Slater, Herb and Doreen Simmers, A. J. Stacey, Wayne Sumner, Alasdair Urquhart, and Grant A. Whatmough.

It will certainly not be thought invidious to give special mention to two people: my doctoral supervisor, James Robert Brown, and my wife, Sharon T. Simmers. This thesis—and indeed my entire graduate career—owes an enormous debt to Jim Brown's continual friendly encouragement and advice. And credit for the completion of the work belongs to Sharon no less than to me.

I could not have carried out this or any other comparable undertaking without the help, either material, spiritual, or intellectual, of certain of the people listed above; they know who they are.

Many thanks are due to the Ontario Centre for Large Scale Computation for the use of computing facilities. I gratefully acknowledge financial support from the University of Toronto and The Social Sciences and Humanities Research Council of Canada through Doctoral Fellowships numbers 452-89-2046 and 452-90-1613.

Note Added in 2019

I have long wanted to generate a cleaned-up, archival version of my PhD thesis of 1991, and here, at last, it is. I have done this in part because my thesis has been cited a few times even though there is no readily available access to it. (It was supposed to have been preserved as microfiche in the thesis repository at Ann Arbor, Michigan, but that is apparently available only to University of Michigan users.) This document will now be archived on OPUS, the open-access institutional repository at the University of Lethbridge (my current home institution), and also on my website.

Unsurprisingly, I now see many ways in which this dissertation could be improved, but I have resisted the temptation to change any part of it except for minor typographical corrections, the excision of a few small infelicities which simply could not be allowed to stand, and some updating of the format. In particular, I have taken advantage of the fact that it is now easy to scan in the graphics, which were originally drawn by hand and cut and pasted into the masters for the thesis literally with scissors and tape. (Fortunately I had preserved my original artwork.) Note that the Table of Contents is now hyperlinked.

After nearly thirty years, how have the major claims of this thesis stood up? Other authors (notably P. J. Bussey, J. B. Kennedy, Jr. and Peter Mittelstaedt) independently arrived at views similar to mine, which is that the usual arguments for peaceful coexistence between relativity and quantum mechanics are question-begging. However, it is safe to say that this critique is not widely accepted. It is also pretty safe to say that it has not been refuted, either. Rather, the deep cognitive dissonance in the received wisdom about the relation between quantum mechanics and orthodox relativity is, for the most part, simply something that one does not discuss in polite company. My own view is that this on-going "nonlocality denial" is one of the major factors contributing to the forty-plus years of utter lack of advancement in our fundamental understanding of physics identified by Lee Smolin.[1] I maintain a stubborn faith that sooner or later physicists, perhaps guided by philosophers of physics who are not afraid to take intellectual risks, will see a way to break the logjam. The keys to the problem are to apply the growing understanding of the dynamics of entangled states to foundational questions, and to accept the shocking possibility that Einstein's Separation Principle may simply be incorrect. I also believe that we need expanded or generalized concepts of simultane-

[1] Lee Smolin, *Troubles With Physics: The Rise of String Theory, the Fall of a Science, and What Comes Next.* Boston & New York: Houghton Mifflin, 2006.

ity that would give us more effective ways to speak of the nonlocal connectedness of quantum mechanics, perhaps along the lines sketched in this thesis.

Some but not all of the ideas in this thesis were developed in publications of mine from 1992 to 2014,[2] and I continue to pursue the challenge of finding a better understanding of nonlocality. I am more than ever convinced that we must accept nonlocality (including *dynamical* nonlocality) as a fundamental feature of the physical universe.

A few omissions in the Acknowledgements I wrote in 1991 must at last be corrected. First, I would like to thank *all* the members of my examining committee. My supervisor was Professor James Robert Brown and my advisors were Professors Alasdair Urquhart and William Seager (all of the Department of Philosophy, University of Toronto). Professor John G. Slater of the Department of Philosophy, University of Toronto, was also on the examining committee, and provided very helpful advice. My ”internal-external” examiner was Professor Jed Z. Buchwald, then of the Institute for the History and Philosophy of Science, University of Toronto, and my external was Professor William Harper, Department of Philosophy, University of Western Ontario. The Chair of the defense committee was Professor H. D. Forbes of the Department of Political Science at the University of Toronto.

Professor Forbes told me that when he (as an expert in political science) saw the title of the thesis he briefly thought that he might be able to understand what was going on, but he abandoned this hope as soon as the *viva voce* got under way.

I am very grateful, as well, to the large community of LaTeX programmers around the world, whose largely unpaid labour has made LaTeX the wonderfully effective instrument for document preparation that it is today.

I should explain the snippet of musical notation that appears under the dedication. My father, Donald Peacock, was a person of many talents, rich imagination, and a great love of life. He was, among other things, a classical pianist, a sculptor, and an architect. The music is the theme of Dad’s signature piece, the Brahms B Minor Rhapsody, Op. 79, No. 1. I have reproduced a screen capture from an anonymous scan taken from the Rafael Joseffy edition, published by Oliver Ditson, 1910 (public domain); it can be found on `imslp.org`.

[2]Most of these papers are collected in my *Quantum Heresies*. London: College Publications, 2018.

Contents

Introduction: Aim and Scope of This Investigation

The object of this thesis is to critically examine the notion of "peaceful coexistence" between relativity and quantum physics. The political slogan "peaceful coexistence"—which refers to a state of mutual tolerance between ideologically opposed camps—was first applied to the relationship between relativity and quantum theory by Abner Shimony (1978). It is meant to suggest that in spite of quantum nonlocality and all the other apparently enormous mathematical and philosophical differences between relativity theory (RT) and quantum mechanics (QM), neither theory (miraculously, one might say) invalidates the other. That is, to say that there is peaceful coexistence between RT and QM, in roughly the sense that Shimony indicates, is to say that there can be no contradiction between the predictions of relativity and quantum theory at the observational or practical level. It is certainly not to deny that one theory may make predictions about problems concerning which the other is silent; for instance, one cannot use relativity alone to calculate energy levels of an atom, nor can one use quantum mechanics to calculate the perihelion advance of Mercury. The doctrine of peaceful coexistence claims simply that RT and QM will never be found to contradict each other in those cases in which they do happen to be talking about the same things—in particular, the types of causal connections that are possible between events. To anticipate, my conclusion will be that no completely satisfactory proof of this proposition has so far been offered. I will also try to suggest that we should not find this surprising—although I cannot offer solace to those who find it disturbing.

The overall spirit in which this thesis is written is a willingness to question that relativity, at least as we presently understand it, really is a so-called "principle" theory—"a set of general constraints which any more detailed theory must satisfy", as Paul Teller puts it.[3] I am quite willing to consider—and in fact I believe it to be probable—that relativity, like all other classical theories, will eventually be understood to be only an approximation to something deeper and stranger. Much discussion will be necessary, of course, to justify this rather iconoclastic position, which goes strongly counter to some

[3]Teller (1989). In fact we have to make a careful distinction between what I will call "textbook relativity" (see section 2.10) and the deeply realistic but yet-to-be-fully-articulated notion behind the Principle of Relativity itself. I think that the latter, if properly understood, might well be salvageable. But that is a topic for another thesis. . .

long-established views about the relationship between relativity and QM.

The problem of the compatibility (or otherwise) between relativity and quantum mechanics is of enormous philosophical and scientific interest in several respects. First, an understanding of the reasons for believing in peaceful coexistence, and of the reasons it breaks down if and when it does, will certainly be essential in consummating that great unfinished project of physics, the long-awaited "fiery union" (in John Wheeler's words) of quantum theory and general relativity. Second, there are fascinating questions about philosophical and scientific methodology which obtrude when we examine more closely the reasons that have been given in favour of peaceful coexistence. Third, and perhaps most important in the long run—especially if we believe that the accuracy, breadth and depth of our understanding of the fabric of the world has a vital bearing on human welfare—there are the deep metaphysical problems that underly the division between these two theories. Relativity is the culmination of the Maxwellian view of the universe, the local realistic theory *par excellence*. Quantum mechanics is a fundamentally nonlocal and arguably nonrealistic theory. Whether or not there is peaceful coexistence in the sense that quantum nonlocality makes no *practical* difference to relativity, we just cannot say that we understand how the world works until we clearly understand how these two radically different views can both have such predictive and explanatory power about the same universe.

We could, without being too inaccurate, think of the doctrine of peaceful coexistence as being a weakened version of the complementarity, or Copenhagen view of the relationship between relativity and QM.[4] In the Copenhagen interpretation of quantum mechanics (which Einstein sarcastically dubbed the "tranquillizing philosophy"[5]) classical and quantal concepts each have their own range of application, and one should not worry about apparent inconsistencies between the two; we can fully expect that the physical world will conform to classical conceptions of causality *where those conceptions can be meaningfully—i.e., operationally—defined.* The problem with this view is that it is now, following the experimental confirmation of the violation of Bell's Inequality,[6] by no means clear that quantum mechanics does not violate the relativistic picture of causality *in the realm in which that picture might properly be expected to apply.* Indeed, an enormous amount of work has been done in recent years to clarify just what notion of locality or causality is in fact violated in "entangled" quantum systems. Currently, the view probably of most workers in this field is that the notion of local realism implicit in the derivation of the Bell Inequalities (and the EPR argument itself) can be factored into two separate conditions, called variously Outcome Independence or Completeness, and Parameter Independence or Locality. (More about the meaning and application of these terms anon.) It is generally believed that of the two only Outcome Independence fails in the face of QM.[7] However, it is quite unclear whether relativity, if properly understood, would actually disallow a violation of Outcome Independence, or whether this is just

[4] This is set forth very clearly in Chapter 23 of Bohm (1951).

[5] Quoted in Fine (1986: 18).

[6] Aspect, Grangier, & Roger (1981), (1982), Aspect, Dalibard, & Roger (1982).

[7] For review and discussion, see Shimony (1986), Ballentine & Jarrett (1987), Redhead (1987), and Jarrett

one of those things about which relativity should be expected to be silent.[8] Therefore the possibility lurks, though few are happy to acknowledge it, that relativity might simply be not merely inapplicable in certain areas, but outright *wrong*. In this climate the classical (pardon the expression) Bohrian notion of complementarity is difficult to maintain.

What is generally believed to save the day for relativity is the existence of several proofs of what we will call the No-Signalling Theorem (NST): the claim that the principles of quantum mechanics alone are sufficient to guarantee that quantum nonlocality cannot be exploited to send readable messages faster than light. (This claim is generally understood to be equivalent to the claim that there can be no violation of Parameter Independence.) Indeed, this is the *only* result that most authors who write on this topic have recently offered as a *guarantee* of peaceful coexistence in the rather limited and precise sense indicated above.[9] That is why the modern doctrine of peaceful coexistence is rather like a much-weakened form of complementarity; it has much the same implications as complementarity (i.e., we can go on using relativity where it would seem to apply) but it hangs only on the slender threat of the No-Signalling Result instead of resting securely on a broad philosophical foundation that would guarantee the unlimited applicability of classical concepts in the classical realm.

Current views are well indicated by the following remarks. First, Shimony himself (1983: 226):

> . . . the non-locality implicit in quantum mechanics must be recognized to be of the uncontrollable variety, that cannot be used for the purpose of sending a signal faster than light. In this sense there may be 'peaceful coexistence' between quantum mechanics and relativity theory.

Other authors echo this view. For instance, Elitzur (1990: 534) says,

> While quantum nonlocality appears to involve an omnipotent influence independent of normal physical restrictions, it is completely useless for any information transfer. This is what maintains the present peaceful coexistence of quantum theory and relativity. Nobody can use the EPR correlations for superluminal communication, because no one can determine at will what result will be obtained in the measurement. One cannot, therefore, affect the other measurement so as to yield the desired 'output'.

(1989).

[8]Jones, Clifton, & Masters (1990) have explicitly argued that even the violation of Outcome Independence invalidates relativity. I have only seen this work in abstract and cannot comment on its cogency. In the following I will show that even violations of Outcome Independence, at least, if not carefully interpreted, seem to open up the possibility of causal looping.

[9]Aharonov & Albert (1981: 369) speak of the way the "covariance of the probabilities" in quantum field theory and the "noncovariance of the state history. . . manage to peacefully coexist"; in their view, this somewhat narrower sense of peaceful coexistence is guaranteed by the locality of all possible observations on any measuring apparatus. In Chap. VII I will discuss the problem of finding a covariant notion of the quantum state.

And d'Espagnat (1976):

> . . . nonseparability does not allow for any *instructions* to be propagated faster than light and therefore, there is no contradiction between nonseparability and relativity, provided Einsteinian causality is conceived of as being of a mere phenomenological nature. . . splitting in two parts the separability assumption that entails the Bell Inequalities. . . [provides] a clear basis for the notion of a peaceful coexistence between relativity and quantum mechanics.

To criticize peaceful coexistence (as it is currently understood) is therefore to criticize the No-Signalling Theorem, and my principal task in this thesis will be to examine the purported proofs of this theorem, and the way they have been interpreted, with some care. I do not want to be thought to be content merely to exacerbate the intellectual crisis occasioned by nonlocality, however, and so toward the end I will offer some constructive, although tentative suggestions as to how we might find our way out of this morass.

It might be useful to to locate this investigation with respect to other foundational problems in QM, by briefly describing what I am *not* going to do in this thesis.

QM is beset by at least three major interrelated interpretational problems, the measurement problem, the problem of nonlocality, and the problem of realism. The problem of measurement is to understand how the apparently discontinuous and irreversible change in a physical system or ensemble of systems upon measurement relates to the continuous and reversible evolution of the system between measurements. The problem of nonlocality is to understand the nature of the apparently instantaneous global change of a system that occurs upon measurement, and to clarify the impact this has upon our relativistic notions of causality. The problem of realism is to understand what QM is actually describing when it talks about physical systems; is the state vector or density operator a property of an individuated reality, or a representation of our possible knowledge of a single real system or ensemble of systems, or something else? The problem of realism cuts across the problems of nonlocality and measurement; for some it is not a problem at all.[10] (It must be said that there are other views as to what is really mysterious in QM. Bohr, of course, thought that the really new thing about QM was complementarity, the existence of mutually exclusive modes of description of the same ineffable reality. Feynman thought that the really mysterious thing was the existence of superposable linear amplitudes as a basis for calculating probabilities; Dirac thought that the really new thing about QM was the occurrence of non-commuting quantities. Like quantum systems themselves, these problems are to some degree nonseparable.)

[10] See van Fraassen (1989).

In the following I will say almost nothing about the realism problem, as serious as it is, except by implication, and will discuss only an important subset of the problems surrounding nonlocality. We shall have no choice but to venture cautiously at least part way onto the minefield of measurement theory. My aim, in a nutshell, is to argue that nonlocality poses far more of a problem for the relativisitic view of causality than is generally presumed in current literature, and my procedure will be to examine in some detail the proofs that have been given for the claim that the non-locality that obtrudes in entangled systems is indeed "uncontrollable". We shall see that if we take a somewhat uncharitable view of these proofs (which to my knowledge have never received anything but sympathetic treatment in the literature) that they are either manifestly circular or else sidestep in some way the very questions they should be addressing. I hope to show, in other words, that the presumed state of peaceful coexistence between QM and RT is, at best, an uneasy truce.

Chapter 1

Some Definitions and Preliminaries

> We shall not here discuss the inexactitude which lurks in the concept of simultaneity of two events at approximately the same place, which can only be removed by an abstraction.
>
> — ALBERT EINSTEIN (in Einstein *et al.* 1950: 39)

1.1 Separability

There has been an enormous amount of effort by many authors in the last few decades to disentangle and define the closely related notions of separability and locality, and many precise definitions, perhaps none of which capture these notions in all generality, have been proposed. *Separability*, in the deepest (or perhaps we should say loosest sense), is a metaphysical concept. The way Einstein thought of it comes out very clearly in these words:

> ...the real factual situation (or state) of the system S_1 is independent of what is done to the system S_2... [Quoted in Fine (1986): 103]

Following Einstein (since it is he more than anyone else who has defined the problems that we are concerned with here) let us say that separability in the deep sense is this:

> Two physical systems are *separable* if they can be separated (i.e., distinguished) with respect to some parameter q in such a way that the real factual state of one system is independent of that of the other.

Note that the parameter q is not necessarily *position*; separability *simpliciter* is a broader concept than separability with respect to space. For instance, one could speak of separability in momentum space as well as ordinary position space. *Locality*, or *localizability*, loosely speaking, is separability with respect to space; I will have more to say about it below.

The problem with this definition of separability is that it is very hard to give it operational meaning; how do we know what the "real factual state" of a system is? (Note that there could also be a problem with making this statement relativistically invariant, since "at a given time" is a frame-dependent concept. This problem will assume a large role later in this investigation.) So we seek a narrower sense of separability that can be expressed in terms of possible sorts of independence of measurable quantities.[1]

Since we are working in a quantum mechanical context, the sorts of separability conditions we will be interested in are probabilistic; they will say that the probability of getting observational results (i.e., of obtaining certain eigenvalues of certain operators) in one system will show some kind of independence of the probability of getting certain other kinds of measurement results in another system. It turns out that there are two probabilistic notions of separability which are of particular importance.

1.2 The Jarrett Conditions

These are separability conditions first described by Jarrett (1984) in his attempt to isolate the premisses that are invalidated by the experimental failure of the Bell Inequalities. Other formulations of his results can be found in Shimony (1986), Ballentine & Jarrett (1987), and Jarrett (1989). For a discussion of their significance in the derivation of the Bell Inequalities, see Appendix A.

There are several ways to formulate these conditions. In simplest terms, a multiparticle system obeys *Parameter Independence* if the probabilities of possible results of measurements carried out on one particle of the system do not depend upon the choice of measurement strategy carried out on other particles in the system. In other words, if Parameter Independence is obeyed, one cannot influence average measurement results on a distant particle by merely choosing to measure a nearby particle in a particular way. Parameter Independence has also been called Jarrett Locality, Simple Locality (Ballentine & Jarrett 1987), Statistical Locality (Redhead 1987), and Surface Locality (Hughes 1989), and Shimony (1984) calls a violation of Parameter Independence Controllable Nonlocality. A system obeys *Outcome Independence* if the probability of a given possible result of a measurement carried out on one particle of the system does not depend upon the measurement *results* obtained on other particles in the system. Outcome Independence has also been called Jarrett Completeness, and Predictive Completeness (Ballentine & Jarrett 1987); Shimony calls a violation of Outcome Independence Uncontrollable Nonlocality.

I will discuss the significance of these conditions in some detail in Chapter 2.[2]

1.3 Phase Entanglement

In quantum mechanical terms, we can also define separability as lack of phase entanglement, or the lack of interference terms, between possible states in the computation of

[1] I do not mean to suggest that the metaphysical notions of separability as independence of real factual states, or independence of intrinsic properties, are meaningless; far from it. I am just pointing out that they are so very difficult to discuss that to make any progress on the question it is necessary to narrow the focus somewhat.

[2] For excellent review and discussion, see the papers by Cushing, Shimony, and Jarrett in Cushing & McMullin (eds.) (1989).

probabilities and expectation values. This is a question that does not arise in classical probabilities, because classical probabilities do not involve amplitudes. This may be the deepest distinction between the classical and the quantum way of thinking, and as Feynman emphasizes[3] we presently have no interpretation or deeper explanation of this strange fact of life. In QM, each possible path a system can take is represented by a complex-valued amplitude. ("Path" does not necessarily mean path through spacetime; it just means a possible sequence of initial, intermediate, and final conditions.) If there are several experimentally indistinguishable outcomes we add the amplitudes *first* and then take the modulus to calculate probabilities.

It is important to realize how fundamentally different this is from classical probabilities. Classically, if $P(e_1)$ and $P(e_2)$ are the probabilities of two mutually exclusive events, then

$$P(e_1 \vee e_2) = P(e_1) + P(e_2). \tag{1.1}$$

Now, suppose we have a QM state $|\Psi\rangle$ such that

$$|\Psi\rangle = \lambda_1 |a_1\rangle + \lambda_2 |a_2\rangle, \qquad |\lambda_1|^2 + |\lambda_2|^2 = 1$$

and suppose β is a possible measurement outcome. The amplitudes $\langle\beta|a_1\rangle$ and $\langle\beta|a_2\rangle$ represent possible "paths" to the final state $|\beta\rangle$. To see that they interfere, in spite of the fact that a_1 and a_2 themselves stand for mutually exclusive measurement outcomes, calculate the probability of getting β in state $|\Psi\rangle$:[4]

$$\begin{aligned} P(\beta) &= \left|\langle\beta|\Psi\rangle\right|^2 \\ &= \left|\lambda_1\langle\beta|a_1\rangle + \lambda_2\langle\beta|a_2\rangle\right|^2 \\ &= |\lambda_1|^2|\langle\beta|a_1\rangle|^2 + |\lambda_2|^2|\langle\beta|a_2\rangle|^2 + 2\mathrm{Re}\{\lambda_1\lambda_2{}^*\langle\beta|a_1\rangle\langle a_2|\beta\rangle\} \end{aligned} \tag{1.2}$$

The third term is the nonclassical interference term; it represents the *entanglement*[5] or interference between $|a_1\rangle$ and $|a_2\rangle$. We can write the weighting factors λ_1 and λ_2 as

$$\lambda_1 = (1/\sqrt{2})e^{i\theta_1}, \qquad \lambda_2 = (1/\sqrt{2})e^{i\theta_2},$$

and this gives

$$\lambda_1\lambda_2{}^* = (1/2)e^{i(\theta_1-\theta_2)}. \tag{1.3}$$

Interference therefore depends in part upon the phase difference or relative phase $(\theta_1 - \theta_2)$.

It is important to realize that this kind of nonseparability can be with respect to any conceivable QM property. For instance, suppose there are two spin-1/2 particles

[3] Feynman *et al.* (1965).

[4] This example is from Cohen-Tannoudji *et al.* (1977: 253–254).

[5] This picturesque term in due to Schrödinger (1935), in Wheeler & Zurek (1983). Hughes (1989b) defines entanglement in terms of the behavior of the tensor product space of a composite system; the non-classical properties he is pointing to are themselves a consequence of interference between possible states.

(which we will call the left and right particle) which interacted in the past but which are now widely separated in space. There are two possible outcomes for spin observations upon the total system: either the left particle is found with spin *up* (i.e. with spin $+\hbar/2$, corresponding to a state $|+\rangle_L$) and the right particle with spin *down* (i.e. with spin $-\hbar/2$, corresponding to a state $|-\rangle_R$) , or *vice versa*. If we knew that the left particle was definitely in the state $|+\rangle_L$, the state of the total system would be given by the tensor product

$$| + -\rangle \equiv |+\rangle_L \otimes |-\rangle_R, \tag{1.4}$$

and correspondingly for the right particle. However, we do not actually get the correct predictions unless we assume that before any measurement is made on the system, its total spin state is

$$|\Psi_S\rangle = (1/\sqrt{2})(| + -\rangle - | - +\rangle), \tag{1.5}$$

the so-called *singlet* state.[6] This is a superposition of *two* states 180° out of phase with each other, and one finds interference terms between these states when one calculates probabilities and expectation values. This means that we get the wrong answers if we try to think of the system as comprised of two independent particles; their spin states are inextricably tangled.

It is a central problem to determine the general conditions under which there is no phase entanglement between possible states of a system of interest—or at least, when the entanglement is small enough to make no observable difference from classical predictions. It may very well be that in the broadest possible sense there are no exactly separable physical systems—only approximately separable systems.

1.4 Local

That which is *local* is that which depends only upon quantities at a point in spacetime, or only upon spacetime position itself. An example of a local quantity would be the air temperature at various points in a room. Clearly, this concept of locality contains an inherent ambiguity connected with the notion of the continuity of spacetime. We know perfectly well that no physical objects are actually infinitesimally close to each other; all objects are actually at a finite distance (in those cases in which we can define a precise boundary to an object at all). And except *possibly* for leptons and quarks, no physical objects are *pointlike*; all objects are extended. (I make the qualification for these elementary particles only because it has not yet been possible to determine in scattering experiments whether they have finite diameters.) Therefore, strictly speaking, it is quite unlikely that there is any such thing as a purely local contact or interaction. However, we believe that we can usually ignore this, or at least hope that we can. Classical physics gets around this puzzle to some degree by attributing the interactions of particles to their associated continuously extended fields; however, this hopeful notion breaks down in modern field theory which requires us to think of fields as themselves collections of particles. The closest thing we have in physics today to any kind of local contact is

[6] See Cohen-Tannoudji *et al.* 1977, Ch. X, for more background.

whatever happens at the vertex of a Feynman diagram, and no one would say that this is well understood!

It must have been, in part at least, the ambiguity in the notion of local contact or coincidence that Einstein was referring to in that tantalizing footnote to his great paper of 1905, which stands at the head of this chapter.

Other senses of "local" are relevant. Streater (1987: 141) tells us that "local" as applied to quantum fields means "defined as averages over arbitrarily small regions". One can also use "local" in a looser sense simply as "nearby", which really just means separated by distances such that the time of light propagation is small in comparison with the time required for other events of interest to take place.

Physicists like local quantities because they are so much easier to understand mathematically. Misner, Thorne and Wheeler (1973: 1200):

> ...general relativity has reinforced the view that 'physics is local'; that the analysis of physics becomes simple when it connects quantities at a given event with quantities at immediately adjacent events.

1.5 Nonlocal

That which is *nonlocal* is that which is dependent not upon or not only upon quantities at a point in spacetime, but instead upon quantities defined over a *region* of spacetime; or, acting throughout a region of spacetime. A homely example of a nonlocal quantity is the *average* air temperature in a room. The notion of nonlocal quantities as such presents no metaphysical difficulties. For instance, it would be quite normal for, say, the cost of heating a room to depend upon the average temperature maintained in that room; there is nothing odd in this. What would be odd would be a case in which a local quantity was determined by a nonlocal quantity. Suppose, for instance (to take an example which we would consider to be obviously absurd), the air pressure at a point in a room depended upon the average temperature, not the local air temperature. The reason this would strike us as absurd is because of the violence it would do to our intuitive notions of causality, which hold that causation occurs only by local contact. For instance, if someone lit a match in one corner of a room, the average temperature would—by *definition*—increase simultaneously with the lighting of the match; however, we would certainly find it bizarre if this could instantaneously affect the air pressure at a different corner of the room, which would be the case if indeed air pressure could depend upon *average* temperature. In fact, however, this is not all that different from what happens in an EPR experiment. A global mathematical object—the wave function—is changed instantaneously by a local operation, the way average temperature is changed by lighting a match at a point in a room. However, this global change of state vector seems to have an effect on certain other local occurrences, namely the measurement results on the other particle in the system.

A local quantity may be treated as a limit of a nonlocal quantity, in cases where a limit can be defined; there are, however, quantum systems in which continuity breaks down and the limit cannot be taken; in such systems certain quantities may be irreducibly nonlocal. For instance, in an atom the energy of each orbital is a property of the orbital as a whole; the assumption that a continuously varying electron energy could be defined as a function of position around the nucleus would produce wildly incorrect predictions in some problems.

1.6 Locality and Nonlocality

I have described Parameter and Outcome Independence as types of separability conditions, even though they are often described in the literature as locality conditions. There is enormous variation of terminology from author to author, and one must draw a line somewhere.[7] I suggest that the concepts of Outcome and Parameter Independence should be called types of separability since they are are cognate to the general notion of separability, which suggests some kind of independence of properties. Were we to reorganize the terminology in this field from scratch, we might use *locality* and *nonlocality* simply to denote the properties of being local or nonlocal respectively. However, these terms have been used in special ways in this context, the broadest sense of which I think might be most clearly rendered as follows: *locality* is separability with respect to, or because of, separation in space and time; (quantum) nonlocality is nonseparability with respect to, or in spite of, separation in space and time. To repeat, there are several possible senses in which systems might be separable or nonseparable; the problem is to determine which senses pertain.

Locality is also often used with the special meaning that there is "no influence transmitted faster than light."[8] The justification for this usage is that if there is arbitrarily fast causation then spatial separation would not guarantee factual independence; therefore certainly in a practical sense this no-superluminal-causation principle is equivalent to locality in my broader sense. However, it would be sterile to quibble about terminology; just be sure to read every author who uses the terms (non)locality or (non)separability carefully, because she or he may be using them in a way you do not expect.

1.7 Principle of Separability (SEP)

The Principle of Separability (Einstein's *Trennungsprinzip*) is the postulate (or perhaps article of faith) that spacelike separation guarantees separability. It was SEP that tripped up Einstein, Podolsky and Rosen (1935), who took it as self-evident that the properties of spatially separate (in effect, spacelike separated) particles cannot be interdependent. Here is the rest of the passage by Einstein which I quoted from in defining separability above:

> But on one supposition we should, in my opinion, absolutely hold fast: the real factual situation (or state) of the system S_1 is independent of what is

[7]Jarrett (1989) is a useful guide to various notions of locality as they are probably most commonly understood these days. Eberhard (1978) gives a very careful and useful spelling out of different possible senses of locality.

[8]Cushing (1989:10), in Cushing & McMullin (1989).

> done to the system S_2, which is spatially separated from the former. [Quoted in Fine (1986): 103]

There seem to be two grounds for the belief in SEP. First, if we believe relativity, we believe that spacelike separated systems are separable because no causal signal can pass between them. There also seems to be a deeper, metaphysical ground, difficult to put into words, but which seems to be a belief that the very notion of existence as a physical object entails existence as a localized, bounded entity in space:

> It is... characteristic of ... physical objects that they are thought of as arranged in a space-time continuum. As essential aspect of this arrangement of things... is that they lay claim, at a certain time, to an existence independent of one another, provided these objects 'are situated in different parts of space'. [Quoted in Fine (1986): 103]

(For a very insightful discussion of Einstein's views, see Fine 1986).

Einstein, like Newton before him, seemed to have taken some sort of Separation Principle to be almost a necessary condition of rational thought. Opinions may differ on how natural an assumption—or prejudice—it really is. But the fact seems to be that SEP, broadly interpreted, is almost certainly false. A study of the derivation of Bell's Inequality shows that what makes the calculation go through is a factorizability condition which follows from the assumption that certain separability conditions hold. Since the Bell Inequalities fail, some or all separability conditions that pertain to the derivation must fail. Thanks to the work of many people we now have some understanding of what goes wrong, but the general problem of determining in what sense and under what circumstances separability of localized objects and events does fail remains to be solved.

Stapp (1989a: 9) states the issue of principle about as clearly as possible:

> The question at issue here is not just a matter of words, or of preference in definitions. Nor is it just an academic philosophical question. For upon its resolution hinges the rational direction of research at the deepest level of science. The question is whether, in efforts to extend our understanding of quantum phenomena from the domain of atomic physics, which is the domain that quantum theory was originally designed to cover, into the domains of cosmology, biology, and mesoscopic physics, one should retain the essentially classical idea that spatial separation entails intrinsic separation, or, instead, shift to a more integrated conception of nature in which the apparent separability of that which is spatially separated is not fundamental, but rather emerges only at some level of approximation.

In quantum mechanical terms, we can probably describe SEP as the assumption that there is no phase entanglement between position states and other particle states (such as spin states).

1.8 Action-at-a-Distance

Action-at-a-distance in the sense we are concerned with here is an instantaneous effect of one body or system localized in space upon another body or system at a distance in space. In relativity theory, any superluminal signal is instantaneous in some frame, and so the relativistic first-signal theorem ("all velocities are finite and bounded by c") is generally taken to be equivalent to a prohibition of action-at-a-distance.

Our present-day worries about action-at-a-distance have evolved from earlier and somewhat different ones. Newton's worry with gravitational action-at-a-distance was not so much its instantaneity but the apparent lack of a medium to convey the effect.[9] Newton, in spite of the fact that he propounded a distant action theory of gravitation,[10] regarded the notion of action-at-a-distance as absurd, and disparaged it in a famous remark to Bentley (quoted in Wheeler, Misner & Thorne: 41):

> That one body may act upon another at a distance through a vacuum, without the mediation of anything else, by and through which their action and force may be conveyed from one to another, is to me so great an absurdity, that I believe no man, who has in philosophical matters a competent faculty of thinking, can ever fall into it.

Einstein, who succeeded brilliantly in transforming Newton's action-at-a-distance theory into a local field theory, shared these sentiments, and spoke of quantum nonlocality as *spukhafte Fernwirkung*.[11]

1.9 Textbook Relativity

By *textbook relativity* I shall mean the more-or-less standard formulation of special relativity that we learn as undergraduates, and in particular the picture of the causal structure of spacetime that obtains in this formulation.[12] I make this slightly tendentious distinction to emphasize the fact that relativity is not a seamless whole. Many authors have questioned whether or not relativity really does prohibit faster-than-light transmission of mass-energy or information, as the textbooks tell us it does; some authors argue that the mathematics of relativity as it stands would permit superluminal signalling if suitably reinterpreted[13] while others suggest revisions to the structure of the theory.[14]

Needless to say, until recently these explorations have not been taken very seriously by most physicists or philosophers of science. Their significance for us is simply

[9] My thanks to J. R. Brown for this observation. For an excellent historical review, see McMullin (1989: 289–290).

[10] Newton is justly credited with many innovations, but one great methodological innovation of his that I think deserves more mention than it has received is his willingness to put to good use a "law" that he knew could only be provisional. This introduced a kind of opportunism to science that it has benefitted from ever since.

[11] "Spooky action-at-a-distance"; quoted in Bernstein (1991: 45).

[12] See Wheeler & Taylor (1963).

[13] Bilaniuk and Sudarshan (1969), Recami (1987).

[14] Sutherland & Shepanski (1986) for instance have published very plausible sets of superluminal transformations based upon the hypothesis that the metric of spacetime is positive *definite.*

that there is sufficient ambiguity in the formulation of textbook relativity to make the consideration of alternate theories possible. But the problem of quantum nonlocality has brought to the fore the possibility that relativity—or at least *textbook* relativity—could be very much in need of repair. Bell himself has argued in several places that the meaning and application of the Principle of Relativity stands in dire need of review. In a recent interview[15] Bell put it this way:

> The theorem [Bell's Theorem] tells you that maybe there must be something happening faster than light, although it pains me to say even that much. The theorem certainly implies that Einstein's concept of space and time, neatly divided up into separate regions by light velocity, is not tenable. [1988: 90]

He charged that we don't really know the answer to the following question:

> What restrictions on velocities—and velocities of what—are really imposed by Lorentz invariance? [1988: 92]

In fact a fairly detailed (although not conclusive) argument can be given that textbook relativity does not actually prohibit superluminal motion, but is merely not formulated to describe it correctly.[16] I think it is useful, therefore, to make a distinction between what relativity as conventionally interpreted does say, and what it might say if only we understood it properly. It is not out of the question that there might be some way of salvaging the Principle of Relativity in its deepest sense, whatever that is, even if Lorentz invariance should be shown to fail in some way. But even if the PR *cannot* be salvaged—so be it; I think that the time has come to openly accept the possibility that the Principle of Relativity is only an approximate symmetry, and let the philosophical chips fall where they may.

1.10 Postulate of Peaceful Coexistence

I have already presented an informal characterization of the notion of peaceful coexistence in the Introduction; however, it is worthwhile to state it once again a little more precisely. Let us say, therefore, that the Postulate of Peaceful Coexistence (PPC) is roughly as follows:

> In spite of quantum nonlocality, and the profoundly different and in many ways more primitive description of the physical world given by quantum mechanics, QM is such as to guarantee that textbook relativity is *exactly* phenomenologically correct.

---

[15] *Omni* Magazine, May 1988.

[16] It would take me far beyond the scope of this thesis to justify this statement, which I realize might strike some people as outrageous. For an argument that relativity does not actually prohibit superluminal motion, see Mundy (1986). For an indication that relativity could suffer from internal problems quite apart from those problems imposed upon it by nonlocality, see Ardavan (1984a,b).

The emphasis on "exactly" in the above definition is essential. If even a *little* bit of superluminal signalling is possible, then—at least on the conventional interpretation—Lorentz invariance would not be an exact symmetry, and we would presumably have lost access to the only universal framework we have within which to formulate physical laws.

This characterization of the Postulate of Peaceful Coexistence (PPC) as a claim that quantum mechanics predicts no *net observable* conflict with textbook relativity (TR) is minimal. In fact, those who advocate the PPC would surely wish it to admit of a stronger formulation. It would presumably be much more satisfying to them if QM did not just *happen* to agree with TR at the level of observable predictions, but if TR and QM agreed in principle in some very fundamental way. (In particular, Shimony (1986) and Penrose (1986) have emphasized the need to find a consistent spacetime description of nonlocal processes in correlated systems.) Whether this means that the two theories should be in some way interderivable, or that one should be derivable from the other, remains unclear. Whatever the ultimate resolution hoped for, however, the main intention is that Lorentz invariance—as conventionally understood—should be preserved in the face of quantum non-locality. We now turn to the question of what it would mean to do this.

Chapter 2

Signalling and the No-Signalling Theorem

> I close these expositions... concerning the interpretation of quantum theory with the reproduction of a conversation which I had with an important theoretical physicist. He: "I am inclined to believe in telepathy." I: "This has probably more to do with physics than with psychology." He: "Yes."
>
> — ALBERT EINSTEIN, in Schilpp 1969: 683.

The question of peaceful coexistence is the question of signalling by the quantum measurement process. In this chapter I will discuss a few matters which will set the stage for our analysis of the proofs of the no-signalling result.

2.1 What is Signalling in General?

Signalling in the broadest sense is the exchange of information between sender and receiver. Information can be defined in a variety of ways, perhaps the most usual being Shannon's definition of the information of a state x as the negentropy of that state: i.e., $-\ln P(x)$, where $P(x)$ is the probability of the state.[1] However, we do not need to concern ourselves very much with how information should be defined, beyond noting two necessary conditions for its transmission.

First, the probability of some observable event or change of state occurring at the receiver's end should be a function of something that happens (I want to leave this as loose as possible) at the sender's end. In other words, the message must be *readable*; the receiver must be able to tell if the interaction occurred or not. Second, the effect must be *controllable by the sender*, although we will see that an uncontrollable interaction can send a message of a degenerate sort.

There may be other requirements for signalling as well—for instance, it could be argued that some sort of exchange of mass-energy between sender and receiver is

[1] See Woodward (1953).

necessary. However, most discussions of quantum signalling have focussed on the questions of readability and controllability, and we now need to define these notions more precisely.

2.2 What Would Quantum Signalling Be?

To understand what quantum signalling would be—or at least what it is widely presumed it would be—we need to define the Jarrett separability conditions a little more precisely.[2] We consider a source emitting pairs of correlated particles L and R in different directions, which we will call left and right for simplicity, and we suppose that at some point each particle is intercepted by some kind of detector which measures some specific quantitative property (not necessarily the same for both particles). Let λ be some sort of specification of the state of the two particles and their source, let p_L and p_R be adjustable detector parameters (such as polarizer angle) which are at the discretion of the left and right experimenters respectively, and let x_L and x_R be possible measurement results (i.e., eigenvalues) on the left and right particles respectively. We assume for simplicity that the possible measurement results take a discrete range of values; in fact, almost any conceivable measurement can be designed so that the outcomes are simply ± 1.

We first define $P_{LR}(x_L, x_R|p_L, p_R)$ as the joint probability that the left device obtains a result x_L and the right device obtains a result x_R when the detector settings are p_L and p_R; this notion of joint probability is considered to be elementary. Let $P_L(x_L|p_L, p_R)$ be the marginal probability that the left device obtains the result x_L when the left and right parameters are set at p_L and p_R respectively and the state of the total system is λ. It is given in terms of elementary joint probabilities by:

$$P_L(x_L|p_L, p_R) = \sum_{x_R} P_{LR}(x_L, x_R|p_L, p_R). \tag{2.1}$$

(For simplicity I won't bother to index these probabilities by λ, but we should assume that they are conditional on a specification of it.) We also define a conditional probability

$$P_L(x_L|p_L, p_R, x_R) = \frac{P_{LR}(x_L, x_R|p_L, p_R)}{P_R(x_R|p_L, p_R)}. \tag{2.2}$$

The marginal probability for a certain left result is the probability of that result together with *any* possible right result. (We assume results on a particular side are mutually exclusive, and use $P(\bigvee_i x_i) = \sum_i P(x_i)$, where the x_i are some mutually exclusive events of interest.) The conditional probability of a certain left result is the probability of that result given that a certain right result was found.

Let $\emptyset$ stand for a null parameter setting, i.e., no measurement was performed. We will say that the system obeys *Parameter Independence* (PI) if the following condition holds:

$$P_L(x_L|p_L, p_R) = P_L(x_L|p_L, \emptyset). \tag{2.3}$$

[2] I am following Jarrett (1989) and Shimony (1986), although with some shortcuts in notation which these authors would probably deplore.

Roughly speaking, this says that the probability of getting a certain left result given that *any* possible measurement was made on the right side is the same as the probability of getting a certain left result given that *no* measurement was made on the right. (Note that the marginal probability in the left side of this equation is defined in terms of a sum over possible right outcomes, not parameters.)

We will say that the system obeys *Outcome Independence* (OI) if the following condition holds:

$$P_L(x_L|p_L, p_R, x_R) = P_L(x_L|p_L, p_R). \tag{2.4}$$

That is, the probability of getting a certain left result conditional on certain parameter choices and a certain right result equals the marginal probability for that result given those parameter choices. Outcome Independence was first called "Completeness" by Jarrett, because if this condition holds the conditional probabilities of measurement results would be a function only of the initial state of the apparatus; the specification of this state would then provide a complete description in the sense that it would provide sufficient information to predict the probabilities of left results. That is, one would not need to know the right results, as well as right parameters and λ, in order to predict left results.

There are a variety of ways of stating these conditions, or conditions roughly equivalent to them. The special significance of Parameter and Outcome Independence as defined here is that, as first shown by Jarrett (1984), the conjunction of Parameter and Outcome Independence implies a factorizability condition which in turn implies the Bell Inequality. (An argument to this effect is given by Shimony 1986: 187–189. See Appendix A.) Since experiment violates the Bell Inequality, either one or both of Outcome or Parameter Independence must fail as well.

Outcome Independence can easily be seen to fail in an "entangled" EPR-Bohm state. For instance, if the Stern-Gerlach detectors are parallel, and I get a spin-up result, I know that (barring detector inefficiencies)[3] my distant partner got a spin-down result. Could we use this for effective signalling?

Suppose the left and right experimenters have made the following arrangement. A pair of particles have been launched from the source in such a way that they will arrive at the detectors at 12:00 noon. The detectors are arranged so that they are parallel to each other. If the left experimenter gets a spin up upon measurement, he will attack by sea; if he gets a spin down, he will attack by land. Whatever result the left experimenter gets, the right experimenter will get the opposite. (As usual, we blithely ignore the possibility of detector inefficiencies.) Therefore, the right experimenter will know which way the left experimenter will attack, quite regardless of how far apart the two are. There is, therefore, instantaneous communication of a rather degenerate sort in an entangled system. This is certainly in itself a striking fact. However, it barely

[3]Some continue to defend local realism by arguing that detector inefficiencies may combine or conspire in such a way as to apparently confirm the violation of the Bell Inequalities. (See Fine (1989), for instance.) I side with those who find this to be extremely implausible; it would be nice, nevertheless, to be able to defeat this position with a quantitative argument.

qualifies as communication, because neither party can choose which message to send; it is a matter of pure chance[4] whether the left experimenter gets spin up or down—and therefore a matter of pure chance which way he attacks, if he sticks to the bargain!

A violation of Outcome Independence in no way guarantees the controllability of the interaction; it simply says that there is a dependency of the probability of the left *result* upon the right *result*; if local results are uncoupled to parameter choices—as they indeed are in the EPR-Bohm device—the decidedly non-classical correlation between left and right results is not subject to local control. For this reason, the violations of Outcome Independence are not generally believed to be a threat to relativistic causality—at least insofar as relativity is taken purely phenomenologically.

I am not completely convinced that we really understand why one could not set up a causal paradox using only a violation of Outcome Independence, in spite of its uncontrollability. Shimony (1986: 192) says, "I see no way of varying the experimental arrangement so as to achieve superluminal transmission" [using a violation of Outcome Independence]. But an inability to think of a way of doing something does not mean that no way of doing that thing exists. However, it would be too much of a digression to explore this very interesting question here. (See Sections 2.4 and 7.8 for related discussion.) We will focus mostly on violations of Parameter Independence, which have been the main concern of most authors, and would be clearly sufficient, at least in principle, to violate TR. But we should keep in mind that Outcome Independence is puzzling, too; eventually, it will emerge that it is not so much controllability of signals, but the connectivity of the worldlines of possible signals, that can threaten relativity.

One way to understand the distinction between Outcome Independence and Parameter Independence is to consider how we might estimate these quantities after we had collected data from a long run of identically-prepared EPR experiments. Suppose Outcome Independence was violated in such a way that the probability of a certain left result given a certain right result was dependent upon choice of right measurement strategy; we could only tell if this had happened by comparing results from both particles. Violations of Outcome Independence show up in *relative* quantities such as correlation coefficients. Data from both sides would have to be brought into conjunction by normal means—i.e., at speeds no faster than the speed of light. (The "communication" scheme outlined above would, of course, only work if both parties knew in advance that Outcome Independence would be violated in a certain way!) On the other hand, Parameter Independence is defined in terms of marginal probabilities which can be estimated (as long-run frequencies) from data from only one particle; hence, if there is anything the right experimenter could do to influence the marginal probabilities found by the left experimenter, the left experimenter could (given a sufficient amount of data) see a variation in left marginal probabilities with left data alone.

Here is how one might be able to exploit a violation of Parameter Independence

[4]Whether this is *necessarily* the case turns out to be a rather difficult question; what we can say, for the moment, is that we presently know of no such experimental arrangement such as I have described here where this would not be the case.

in order to achieve controllable superluminal communication. Suppose there was some parameter value p' for which the condition failed. The left and right experimenters can communicate by making the following arrangement. A source of particles is prepared which emits a long sequence of pairs of particles. The left experimenter agrees to set his detector at a fixed parameter value and compile results as particle after particle impinges upon his detector. The right experimenter switches his detector between p' and some other parameter value (or alternates between measuring at parameter p' and making no measurement at all) in some meaningful pattern (say Morse Code); *if* PI is indeed violated, the statistics of the left experimenter's results will vary in step with the variations in the right experimenter's measurement strategy. If the efficiency is low the right experimenter might have to make a large number of measurements to get one "bit" through, but with patience he will succeed. Therefore it looks like we could use a violation of Parameter Independence to achieve quantum signalling. (I discuss a more detailed scheme in Section 4.3).

2.3 The No-Signalling Theorem

From the argument above, we see that if Parameter Independence is ever violated then relativity is violated. Fortunately, though, says Shimony (1986: 191):

> The quantum mechanical predictions concerning ensembles of pairs of particles do not violate Parameter Independence, provided that non-locality is not explicitly built into the interaction Hamiltonian of the particle pair. Specifically, it is impossible to capitalize upon the entanglement of the quantum state of a two-particle system for the purpose of sending a message to an observer of one of the particles by performing an operation upon the other particle. This is a theorem which has been proven in great generality by Eberhard (1978), Ghirardi *et al.* (1980), and Page (1982)...

Let us call this claim the No-Signalling Theorem (NST), and state it as follows:

> **No-Signalling Theorem:** No quantum-mechanical system violates Parameter Independence. That is, the probability of obtaining a given experimental result in one branch of a correlated system is independent of the *average* results obtained in the remote branch, and even independent of whether or not a measurement was performed on the remote branch.

We will take the NST to be a necessary and sufficient condition for peaceful coexistence. Necessary, because if it fails textbook relativity fails; sufficient, because *(a)* peaceful coexistence is content if relativity is not violated phenomenologically, *(b)* we assume that violations of Outcome Independence do not violate relativity phenomenologically, and *(c)* no one has thought of any other way of violating relativity by means of quantum or particle physics (apart from the possibility of tachyons, which are at best a speculative possibility). Therefore our task will be to examine the proofs that have been offered of the NST and see if they really do what they are supposed to do.

2.4 Why Worry About Signalling?

There are at least three reasons why the possibility of superluminal signalling is felt to be a threat to the whole structure of modern physics, if not to the rationality of science itself:

1. Some believe that it threatens relativity as a principle theory.
2. It seems to allow the possibility of causal anomalies.
3. It amounts to a kind of action-at-a-distance, which is viewed with deep abhorrence by many people.

I have already mentioned the possibility that quantum nonlocality, and in particular the possibility of signalling *via* quantum nonlocality, poses a grave threat to the special status of relativity as a principle theory. Pitowsky, discussing the possibility of signalling in a hidden variable theory, says (1989: 46):

> ... any *theory*, based on hidden variable states... could not be made covariant. If we take such a theory to be true, then we reduce special relativity to the status of a statistical rather than a principle theory of space and time.

Whether the reduction of relativity to a *phenomenologically adequate* statistical theory would really threaten its status as a principle theory is perhaps debatable. It certainly would to the extent that "statistical" means "only approximately true", and whether or not this is the case is our main concern here.[5]

Why would the breakdown of Lorentz invariance as an exact symmetry be such a problem? The opinion of probably most physicists is expressed by Kaufmann (1977: 28–29):

> The entire framework of modern physics is intimately related to consequences of special relativity. These consequences dictate that the speed of light is a barrier that cannot be exceeded under any circumstances.

Certainly many theorists feel that the loss of the "security blanket" of Lorentz covariance would be a grave blow to our ability to formulate useful new theories. Whether this is correct is hard to say. Nevertheless, symmetry and invariance principles have been of the greatest utility in the progress of physical science, and I think that people would be far more willing to countenance the possibility that relativity is an approximate symmetry if we had any idea what deeper symmetry principle, encompassing what we now understand of relativistic and quantal phenomena and possibly including other things as well, could replace it.

[5]Einstein seems to have thought of "principle" theories as those theories which are so well established by experience as to not be in any way conjectural. He contrasted them with what he called "constructive theories", which were fallible, explanatory hypotheses. Whether the distinction is really cogent, is beyond the scope of this thesis. For a sympathetic discussion of Einstein's view of "principle" versus "constructive" theories, see Brown (1991); for Einstein's original statement of the distinction, see Einstein (1919).

Probably the strongest justification for the concern about the loss of Lorentz covariance as a guiding principle is the possibility that a theory that could not be made Lorentz covariant might threaten causality; and indeed, the most serious problem associated with superluminal signalling is the apparent possibility of causal anomalies.

There are two types of causal anomalies that could occur, or so it seems, if superluminal signalling were possible. First are the violations of the order of cause and effect that would occur if a signal could be transmitted over a one-way path between two spacelike separated points. Any two spacelike separated points will differ in temporal order in different frames; if a message was sent from one to another in one frame, there would exist another frame in which it would appear that the reception occured before the transmission. However, it is not clear that this would not be anything more by itself than an *odd effect*, not fundamentally any different from the fact that sounds emitted by a supersonic aircraft can be heard in reverse order by a suitably placed observer.

Far more serious are the paradoxical situations that seem to be possible if superluminal messages could be sent around a closed loop in spacetime. If a superluminal message can be sent from one observer to another who is moving away from the first with sufficient velocity, a return superluminal message can intersect the world line of the first observer at an earlier proper time than when the first message was sent. (This is a result which follows strictly from the equivalence of all Lorentz frames.) If messages can be send back in time in this fashion, creating a so-called closed causal loop, it appears that logical paradoxes of a particularly vicious sort could be set up. It is easy to show that without some sort of global restrictions on the types of messages that can be sent around such loops, it is possible to set up scenarios in which a certain event occurs if and only if it does not occur. To see how this works, suppose there is some sort of signalling device which sends signals which are always instantaneous in the rest frame of the sender. (See Figure 2.1.) Let the lines $x = 0$ and $x' = 0$ represent the worldlines of two experimenters who are moving away from each other. An experimenter at world point O sends an instantaneous message to world points O'. The worldline of this message lies in the plane $t = 0$, the plane of simultaneity in O's frame. The receiver at O' immediately replies. The worldline of this return message is the plane $t' = 0$, which intersects O's worldline at A, which is at an earlier proper time than O. Now to set up a paradox, suppose that the device in the left laboratory has a switch which turns the device off if and only if a signal is detected at A; suppose that the device will certainly operate if this switch is not turned off; and suppose the device in the right laboratory sends a signal at O' if and only if it detects a signal at O'. Then if the message is sent at O the device gets turned off at A and the message is not sent at O; on the other hand, if the message is not sent at O then the device does not get turned off and the message is sent at O. Therefore, the transmission event at O occurs if and only if it does not occur.

Paradoxes like this have been debated endlessly in the literature. Although there seems to be something suspicious about the whole line of reasoning, no universally convincing way out has been found that does not involve setting up some sort of preferred frame of reference or global concept of simultaneity, which most authors

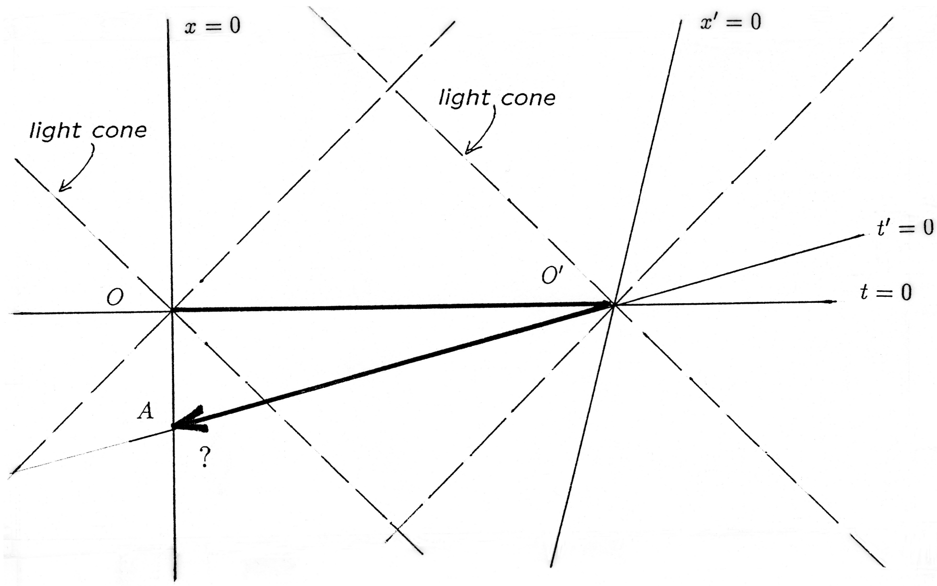

Figure 2.1: A closed causal loop.

believe would be a blatant violation of the Principle of Relativity. (In Chapter 7 I shall argue that this is probably the very position that nonlocality forces us into.) Eberhard (1978: 403) echoes the general view: if quantum signalling were possible then we could "...send information to any point of the past. [This would] lead to severe difficulties in our concept of the Universe." Presumably these difficulties would be either that the universe is fundamentally irrational at some level (as it would be, if we could not find a way of describing it which did not force us to say that there are events which could occur iff they did not occur) or that Lorentz invariance breaks down.

A key question is whether paradoxes of this nature depend upon the superluminal signal being *controllable*. Recall that we have noted a distinction between controllable nonlocality (violations of Parameter Independence) and uncontrollable nonlocality (violations of Outcome Independence). Jarrett (1989: 77) says:

> ...no correlations of the sort associated with violations of completeness [Outcome Independence] can be exploited for superluminal communication because it is a consequence of the failure of determinism that measurement outcomes are not (even in principle) under the control of experimenters.
>
> Incompleteness, then, appears to represent a connectedness of some sort between spatially distant events...which nevertheless does not directly contradict relativity. Abner Shimony has introduced the term 'passion-at-a-distance' to refer to this subtle form of nonlocality. In principle, it is only the uncontrollability resulting from indeterminism that stands in the way of superluminal communication via passion-at-a-distance.

It may very well be true that we cannot exploit violations of Outcome Independence for communication in the sense of being able to send information, but it is not clear that even this sort of nonlocality does not "directly contradict relativity". What really seems to make the above paradox work is just that the loop can be set up at all. There seems to be no reason why we could not design the device at O' to transmit if and only if it detects *any* signal at all, regardless of what that signal is. Similarly, there seems to be no reason why we could not design the switch at A to shut down the left transmitter if and only if it detects any signal at all, regardless of what that signal is. The controllability or otherwise of the signals seems to be quite irrelevant to whether or not a paradoxical loop is possible; hence violations of Outcome Independence could be much more of a problem for relativity than is usually supposed. In a later chapter I will argue that the simplest way out of this might be to deny that EPR-like correlations are established instantaneously in every frame; in other words, I will argue that quantum nonlocality defines an invariant simultaneity-like relation in spacetime.[6] For now, though, let us take it that the problem to be concerned with is "controllable" quantum nonlocality, which we will broadly interpret as any kind of nonseparability in which measurement results in one wing of an extended system shows a dependency upon the *choice* of measurement procedure in the other wing.

Another reason that the notion of superluminal signalling is viewed with deepest suspicion is that any signal outside the light cone is instantaneous in some frame and would therefore amount to a kind of action-at-a-distance, a notion toward which we feel an almost instinctive distrust. It must be admitted that the very notion of instantaneous action-at-a-distance seems somehow contradictory, since we intuitively think of action, interaction, or causation as some sort of process that takes time, that proceeds *from* a point, object or event *to* another point, object or event, *via* some sort of intervening medium. I am not personally convinced that the notion of action-at-a-distance really is absurd; the fact that it offends our mechanical instincts could be just an anthropocentric prejudice based on the fact that so much of our bodily experience is of direct physical contact with our surroundings. The real problem with this spooky quantal action-at-a-distance is our present inability to describe it relativistically without it appearing to be *logically* contradictory. In the concluding chapter I will suggest that further analysis of the concept of simultaneity may give us a more productive way of thinking about the apparent action-at-a-distance in quantum systems.

2.5 Why the NST Seems Reasonable

A large part of the purpose of this investigation is to cast doubt upon either the NST itself or certain conclusions that have been drawn from it. However, to put this in proper perspective, I should take some trouble to explain why it is indeed reasonable to believe in the no-signalling property of quantum systems, above and beyond any of the formal grounds for this theorem—*and* apart from the fact that signalling would

[6] This, of course, certainly contradicts relativity, or at least textbook relativity. However, I prefer a contradiction to TR than a contradiction to logic, which is what we get if we believe in closed causal loops.

apparently cause such problems for TR.

The greatest *practical* barrier to quantum signalling, of course, is that the results of *individual* quantum measurements are completely uncontrollable. What an observer at one of the detectors in an EPR-Bohm apparatus actually sees is an apparently *random* sequence of ups and downs. No variation of the detector parameters or any other kind of experimental manipulation that we presently know of can allow the observer to *choose* to get (say) spin *up* for a particular particle. From the local observer's point of view the result of an individual measurement is entirely a matter of chance. Therefore, in practical terms, sending a message with an EPR apparatus would be like trying to send Morse code without knowing if the key would send a dot or a dash; little help it is for communication to know that whatever the key does chose to send, it is echoed at the other end.

In general, as far as we know, all individual events in quantum systems are entirely random; patterns appear only in the averages. At risk of belaboring the obvious, we should note that this is not known to us *a priori*, but as a generalization from experience. Jarrett (quoted above) stated that measurement results "are not (even in principle) under the control of experimenters". Perhaps; but it is quite unclear what "in principle" means. If this means anything in the context of this discussion, it means "according to the principles of quantum mechanics". However, I suspect that one could make a very good case that the mathematics of QM does not actually *establish* the local uncontrollability of measurement results, but implicitly *assumes* it. QM could perhaps even be defined as the mathematics of random nonlocal events. (In fact, this is the sort of picture that will emerge in a later chapter.) In any case, I am not concerned to deny a very reasonable statement which is a generalization from a very considerable amount of experience with QM systems of all sorts; I simply insist that we be clear on what follows from what.

Granting this fact of randomness, however, the only way that superluminal signalling might have a hope of succeeding in the face of quantum randomness would be if the *average* results obtained at one branch of the apparatus could be affected by detector "twiddling"—deliberate variation of some detector parameter—at the other branch.

To get some feeling as to whether this might be possible, consider again the singlet state, $|\Psi_S\rangle = 1/\sqrt{2}(|+-\rangle - |-+\rangle)$. It is well known that in this state the correlation between results obtained at detectors set at a relative angle θ is $-\cos\theta$. However, the probability of obtaining one of the two possible measurement results at a given detector is just $1/2$.[7] This expresses the fact, mentioned earlier, that the average results obtained by an observer at one branch of the apparatus are completely random, depending in no way on the setting at the other branch; non-locality is exhibited only by a *relative* quantity such as the correlation between measurements at the two branches, which depends upon some relation between the properties of the apparatus at both branches. Although the correlation is nonclassical, it is only seen to be so when we have both sets of results to compare, and these results must be brought together for comparison by

[7] See Appendix B.

ordinary means. This is just another way of saying that Parameter Independence holds for the singlet state, while Outcome Independence fails; "you can send a message, but you can't decode it!"[8]

In the singlet state, therefore, even the hope of signalling by controlling the long-run statistics of results is out of the question. There seems to be no way in which acts of measurement at one branch could be used to send a readable message to the other branch, regardless of how much detector twiddling one does. This is known to be true for many familiar systems such as the singlet state; *the aim of the NST is to show that this is true in general.* The important point here is that the local uncontrollability or randomness of quantum measurement results is not merely theoretical but is well-confirmed by very broad experience with quantum systems of many kinds. The condition described by the statement of Parameter Independence is an expression of this fact of experience. Now, the question arises as to whether there are any more general grounds from which we can somehow prove that quantum systems will always obey this condition. The problem we shall be faced with is whether the mathematics of quantum theory can establish the local uncontrollability of measurement results, or somehow *assumes* this implicitly.

Another argument for the reasonableness of the NST has to do with certain thermodynamic peculiarities of nonlocal interactions, emphasized in a recent paper by Elitzur (1990). As Elitzur points out, there is no energy exchange between the distant particles in an EPR apparatus. In other words, if there could be signalling simply by the measurement process, it would be unlike any other signalling that we know of in that it would be thermodynamically *free*. (This is not strictly true, since a small amount of energy must be expended locally to adjust the detector parameters.) Signalling or exchange of information does not necessarily involve transmission of mass-energy *to* a remote site; it can involve removal of mass-energy *from* a site; however, it does always seem to involve exchange of energy *between* sites (in a detectably improbable pattern). This factor by itself makes the whole idea of sending messages over arbitrary distances seem extremely unlikely—at least, if we envision that it could be done merely by twiddling a polarizer or Stern-Gerlach detector.

There is yet another type of argument (admittedly very imprecise) for the reasonableness of the NST which does not, so far, seem to have been considered in the literature.

Bohm (1980) has pointed out that we could perhaps construct a realistic model for quantum phenomena along something like the following lines (which I will call the *projective* interpretation of quantum mechanics, since it does not yet seem to have been given a name). Let us suppose (without worrying too much about making this image precise) that all physical processes really happen not in spacetime but in a sort of central *non-spatially extended* focal point, monad, or kernel; if this means anything, then the two correlated particles in an EPR apparatus are not really two separate particles, but one and the same object somehow projected onto two different locations in ordinary

[8]This phrase is due to Herbert (1982).

spacetime the way that the shadows of a single object might be projected onto two screens at right angles to one another. This simple but sketchy model easily reproduces two of the most important features of EPR-like systems. First, whatever happens to the central object is, of course, mirrored in perfect step (or at least in some definite phase relationship) in the two projected images, so it could appear that there is some sort of action-at-a-distance or instantaneous communication between the images. Second, this apparent "interaction" between the images is quite uncontrollable by an experimenter located near either of the images, because whatever changes the images undergo is obviously not determined *at the site of the images* but at the central location, wherever that might be. Although vague, this analogy is tantalizingly suggestive, and by placing the "true" object literally outside ordinary spacetime it might offer the last hope of realism (if we think of realism as a program which seeks to interpret the quantum object as a discrete and independently existent entity). If there is anything to a model of this nature at all, then the local uncontrollability of nonlocal correlations is entirely natural. I emphasize, however, that these remarks must be accounted a speculative aside.

The NST is indeed an extremely plausible generalization from our broad experience with quantum systems and in particular from the fact of experience that local measurement results are uncontrollable. But the question of the truth of the NST is, of course, logically distinct from the question of the adequacy of the proofs that have so far been offered for it. We will now consider some of the problems that arise when one attempts to construct a completely general proof of this apparently unobjectionable claim.

2.6 Defining the Logical Problem

In this chapter I will try to clarify the logical problem that I believe we are faced with in trying to defend the Principle of Peaceful Coexistence. My principal complaint against the NST proofs that have so far appeared in the literature is that they are all, to a greater or lesser degree, *circular*. However, it is very easy to misunderstand what I mean by this, and a slight digression into some basics of informal logic would probably be helpful.

A circular argument is an argument which includes among its premises the very claim that it purports to establish. It therefore has the general form

$$A,\, B \vdash A \tag{2.5}$$

where A is the proposition we would like to establish, B is the conjunction of various other premises which we will call the *background conditions,* and where the "$\vdash$" sign can be read as "therefore". Let us say that an argument of the form (2.3) possesses *Simple Circularity*. Such an argument form is, of course, trivially deductively valid,[9] but is fallacious in the sense that it most certainly fails to establish the claim (or more precisely

[9]An argument is deductively valid if and only if the conclusion follows from the premisses by correct application of the rules of logical inference; this is the same thing as saying that the premisses of the argument could not be true and the conclusion false. An argument $A \vdash B$ is deductively valid if and only if the conditional $A \rightarrow B$ is a logical truth.

the *metaclaim*) that A follows from *B alone*. To argue in a circle is sometimes also called "begging the question" or *petitio principii*.[10] Question-begging is one of several *fallacies of relevance*, so-called because the premises of such arguments are, as Copi (1982: 107–108) puts it, "logically irrelevant to the purpose of *proving* or *establishing* their conclusion." Copi (*loc. cit.*) explains very succinctly why it is fallacious to argue in a circle:

> If the proposition [the purported conclusion] is acceptable without argument, no argument is needed to establish it; and if the proposition is not acceptable without argument, then no argument that requires its acceptance as a premise could possibly lead one to accept its conclusion...

It is surprisingly easy to fall into the trap of begging the question, since the circularity can be hidden from us in a number of ways. We might, for instance, put forth an argument of the form

$$A', B \vdash A \tag{2.6}$$

where A' is a proposition which is either logically equivalent to A, or else merely the same statement as A but in different words. (Let's call this *Hidden Circularity.*) It might turn out to be quite difficult to show that $A \iff A'$. Or else we might simply *forget* that at an earlier time we already adopted the proposition we are trying to prove; this might lead to an argument of the form

$$(C), (C \to A), A, A \to C \vdash C, \tag{2.7}$$

where bracketing indicates forgotten assumptions. (Let's call this the *Fallacy of the Forgotten Premise*.) Again, this is a surprisingly easy trap to fall into. Here is one way that this might happen: suppose I am trying to establish a claim within a rather extensive, if not sprawling field of study (such as, say, quantum field theory) and I fail to pay sufficient attention to my historical antecedents; I fail to see that anyone who works in my field implicitly accepts the proposition I am trying to prove and in fact uses it to ground the other propositions that I am trying to derive my claim from. So even though my argument in isolation might not be circular, it forms a fragment of a circular argument if placed in the context of the whole field I work in. We will see that something very much like this seems to have occurred in the quantum field theoretical proofs of the NST.

There is another version of the Forgotten Premise which is not uncommon. We argue

$$(A'), B \vdash A, \qquad \text{where } A' \iff A, \tag{2.8}$$

and then when someone points out that we have smuggled in the premise A', we say, "Well, look, A' (or A) is a reasonable assumption *anyway*."

[10] It is becoming increasingly common for the phrase "begging the question" to be used loosely with the sense of "evading the question". This is a regrettable (but all too typical) blurring of meaning of a very useful expression. Question-begging is of course a kind of evasion, but it has the precise meaning that one is assuming, perhaps covertly, implicitly, or unwittingly, the thing that one claims to prove.

Another practice closely related to question-begging, if not actually identical to it at times, is *Defining a Problem out of Existence.* Suppose a long-established theory contains the following theorem, which happens to be well-confirmed by experience: "No A are B." Suppose also that it is becoming evident that there are some A's of a rather deviant sort that *are* B. Adherents of the theory, rather than see their favourite theory lose its absolute status, may say things like: "Well, *those* A's aren't what we mean by an A, and anyway they are so rare that we can ignore them"; "Yes, but we won't *call* it a B in this case"; "That's embarrassing, let's not talk about it"; or even "This phenomenon couldn't be real because we don't have a theory to describe it". This sort of talk is often heard as a theory becomes increasingly scholastic and self-confirmatory.[11]

It is easy to mock these practices. However, the tendency toward circular thinking runs very deep; if my worldview includes a certain proposition, it may be extraordinarily difficult for me to defend that proposition without somehow assuming its truth; I might not be able to even formulate a description of the grounds on which it could be based without implicitly using that very proposition. I am not trying to make a Quinean sort of claim here that all human knowledge is unavoidably circular. (In fact, I think that circularity is something we should avoid; at the very least, we should try to make the circle as large as possible.) I am just saying that as a matter of sheer practical fact it can be very hard to sort out what follows from what. This in no way absolves us of the responsibility to do so, however.

2.7 The Problem: Can We Prove NST from QM Alone?

The problem of circularity is particularly acute in evaluating the notion of peaceful coexistence, because it can and has been argued that QM and classical physics (including relativity) necessarily stand in a circular relationship. This is a possible interpretation (or misinterpretation) of the influential Copenhagen doctrine of complementarity, according to which quantum and classical concepts, even though mutually inconsistent if applied to the same classes of phenomena, each have their own proper sphere of application within which they are perfectly correct and logically consistent. I have already mentioned that quantum nonlocality threatens this neat partitioning of physical theory; however, we need to understand how persuasive this neo-Kantian doctrine is.

The basis of the notion of complementarity is the entirely correct observation that we cannot formulate results of possible measurements, or more generally describe any definite conditions of a physical system, without the use of classical language. The only possible states of a particle that we can know about are states that express themselves macroscopically, and the only way we know how to formulate a problem is in terms of macroscopically accessible base states and Hamiltonians. Thus, in books that take a strong Copenhagen line, one will find statements such as these:

> . . . quantum mechanics occupies a very unusual place among physical theories: it contains classical mechanics as a limiting case, yet at the same time

[11] See Lakatos' (1976) discussion of "monster-barring" in mathematics.

> it requires this limiting case for its own formulation... [Landau & Lifshitz 1977:3; quoted in Bell 1990:35.]

This careful statement seems to be quite correct. However, we find Bohm saying this:

> ...quantum theory presupposes the classical level and the general correctness of classical concepts in describing this level... [Bohm 1951: 625][12]

The problem is that there is a very significant ambiguity in Bohm's expression "correctness of classical concepts". It is certainly true that any concepts we use to represent possible conditions observable by human beings have to be in terms of concepts which are meaningful at the human level; perhaps these could be called "classical", but there is a danger here that we will get confused between the obvious necessity for our description of the world we can observe to be couched in terms of human-scale concepts, and the idea that the description of this world must conform to that set of laws which for historical reasons have come to be called "classical". This is the ambiguity I mentioned: the *applicability* of classical concepts (say force, electric charge, momentum, etc.) does not imply the unconditional truth[13] of those lawlike relationships between these concepts that we presently happen to be familiar with. The fact that QM requires classical language for its formulation in no way implies that QM can be expected to confirm classical laws as we presently understand them.

There is a possible objection to what I have said here: it might be said that if the meaning of classical concepts such as force, momentum, energy, etc., are precisely specified then certain relationships (which we might call the laws of classical physics) automatically hold between them; hence to use these concepts is necessarily to hold to certain laws in the classical realm. However, I think this is to take the formal definitions of terms as they appear in theories too seriously. The definitions of concepts as they appear in theories often get narrowed in order that those concepts may obey precise laws; in fact, "classical concepts", being simply names for effects and regularities observable on the human scale, are, and ought to be, somewhat loose and subject to revision. (This is a point that Bohr emphasized.) There is obviously much historical precedent for this: witness the way many classical concepts, such as force, momentum, mass, and so forth, were redefined with the advent of relativity, or the impact that nonlinear dynamics has had on classical physics in the last two or three decades.

I see absolutely no reason to suppose that the set of concepts we now (1991 AD) call "classical", or the relationships that hold between them, are fixed for all time. Therefore I see no reason to suppose that simply because the only quantum states we happen to be able to define are those that correspond to humanly observable macrostates (through what Wheeler calls an "irreversible process of amplification"), that the regularities we

[12] In the same passage Bohm also makes the odd statement that QM "does not deduce classical concepts as limiting cases of quantum concepts." What Landau and Lifshitz say seems to be far more accurate.

[13] *Truth* and *correctness* are significantly different concepts which are often conflated. Correctness is adherence to formal rules; truth is that hard-to-define quality that our representations have when they do somehow, to some recognizable degree, represent what whey are intended to represent.

now are familiar with between those macrostates are the only regularities that will ever become apparent, or that we might not even be able, some day, to identify other regularities and even other macrostates of a type with which we are not now familiar.

In short, I see no reason why we cannot expect QM to lead us beyond classical mechanics as we presently understand it; and I especially do not see that we should expect QM and classical physics to always be mutually confirmatory. This means that the odd circular relationship between quantum and macroscopic concepts that Landau and Lifshitz point to gives us no warrant to borrow a principle that is, broadly speaking, classical, and try to conjoin it with some principles of QM in such a way as to justify that very same classical principle. A circular argument in this context is quite as fallacious as in any other. In any case, whether or not we think that this circularity is unavoidable (and some will still think this way) we have to acknowledge that it is there.

I have taken such pains to spell this out because the claim has been made by several authors that their proofs of the NST are proofs *from the principles of QM alone.* Indeed, no one would think that these arguments would support the PPC at all if this were not the case. The whole point of the inquiry is to see whether or not quantum mechanics *without any special side conditions* confirms the relativistic picture of causation.

Here are some representative statements from the no-signalling literature:

> ...quantum theory has property 4 [Parameter Independence]...[Eberhard 1978: 416][14]
>
> ...quantum mechanics satisfies...the simple locality that is needed to satisfy the demands of relativity...[Ballentine & Jarrett 1987: 696]
>
> ...some authors claim that EPR-type setups...would allow, through wave-packet reduction, faster-than-light communication...That this conclusion is certainly false is a very simple consequence of quantum rules...[Ghirardi 1988: 95]

Now we will see if these authors have indeed accomplished what they say they have.

2.8 Introduction to the Proofs

There are two distinct classes of proofs of the NST that have appeared so far in the literature. One relies upon a principle of quantum field theory (QFT) variously called microcausality or local commutativity; the other class of proofs can be loosely described as following from the measurement theory of non-relativistic QM, although in fact they essentially all depend upon the Separation Principle, SEP, in some form. What makes both types of proof work mathematically is the implicit or explicit commutation of operators associated in some way with the spacelike separated branches of an entangled system, or upon some condition essentially equivalent to this; where these proofs

[14] Eberhard has been much more precise, in recent years, about the scope of his proof; see Eberhard & Ross 1989.

methods differ is in the way they justify this commutation. The overall strategy is to show that there is no dependence of the probability or expectation value of a certain result on the left upon the measurement procedure on the right. Usually, we will end up showing that the left results will be the same whether or not any measurement was made at all on the right.

Our concern here is with attempts to prove the NST as a general property of QM, and it is only proofs of this type of no-signalling result that we will consider in detail. We should note, however, that there have been a number of papers published which argue against specific signalling schemes that have been proposed.[15] Also, Troudet (1985) and van Kampen (1984) have presented arguments to show that the Bohm-Aharonov Effect cannot be used for superluminal signalling. Troudet's proof uses Feynman path integrals; I am not aware that anyone has attempted to prove the NST in all generality using this method. It might be very instructive to do so, because the path integral formulation of QM is quite distinct from, and possibly of greater generality than other ways of doing quantum mechanics.

In the next three chapters I will set forth and criticize typical examples of both of these kinds of proofs and try to clarify the logical relation between them; I will argue that these proofs, while instructive in many ways, offer no support for the PPC as a categorical proposition.

[15] Ghirardi & Weber (1983), for instance, refute Herbert's (1982) FLASH scheme.

Chapter 3

Proofs from Non-Relativistic QM

> ...we have considered worthwhile to illustrate explicitly the general proof of the impossibility of superluminal transmission, even though it is quite elementary, to stop useless debates on this subject.
>
> — G. C. GHIRARDI, A. RIMINI, & T. WEBER (1980: 298)

> It is one of the chief merits of proofs that they instil a certain scepticism as to the result proved.
>
> — BERTRAND RUSSELL (1903, quoted in Lakatos 1976: 48)

Most published proofs of the NST are done within the framework of non-relativistic quantum mechanics (NRQM). (Here is a representative sample: Ghirardi, Rimini & Weber (1980), Bussey (1982), Page (1982), Jordan (1983), Shimony (1983), d'Espagnat (1984), Redhead (1987), Ballentine & Jarrett (1987), and Eberhard & Ross (1989).) These NRQM proofs are somewhat difficult to characterize and classify, both because a great variety of mathematical methods have been used and because some of the authors do not spell out the physical justification for what they do as clearly as might be hoped. However, as far as I can tell, they really all come down to much the same thing, which is to work out the consequences of an implicit or explicit acceptance of some notion of the separability of two subsystems of a spatially extended, entangled system, based upon their presumed *dynamic independence*; this assumption is just cashed out mathematically in different ways by different authors. In this chapter I will sketch a number of well-known NRQM proofs with the primary intention of showing their reliance upon this notion of separability.[1]

Recall what these proofs aim to establish: even though there are nonclassical correlations between nearby and distant measurement results (violations of Outcome

[1] In this and the following chapter I have profited greatly from "Quantum Field Theory Cannot Provide Faster-Than-Light Communication", by P. H. Eberhard and R. R. Ross (1989). This paper is by far the most comprehensive review of the signalling question published to date.

Independence), nearby measurement results are completely statistically independent of *distant* parameters—or more broadly speaking, distant choices of measurement strategy (this is what we have called Parameter Independence). The physical reason for this is that nearby results are statistically independent of *nearby* parameters. (This is what we mean by saying that they are “random”.) Now, the question is how we are to establish, or perhaps merely express, this fact of life formally.

I should be careful to point out that nothing I am going to say in this chapter is unknown to physicists, or should even be considered surprising. My purpose is simply to serve a reminder that certain long-established ways of thinking about quantum systems have a certain significance for the discussion of peaceful coexistence.

3.1 Jordan’s Proof

The prototype of all NRQM proofs, and by far the most succinct of all proofs of the NST, is that offered by T. F. Jordan (1983). I will reproduce his argument here, spelling out the steps in somewhat greater detail than he does.

Referring to an earlier paper by Bussey (1982), which proves a no-signalling result for the case of two correlated spin-1/2 particles, Jordan states that “quantum mechanics predicts” that the non-classical correlations in entangled systems such as the singlet state “cannot be used to send signals”. He then sets out to show that “this is a general property of correlations between separate subsystems of any quantum-mechanical system. . . ” I have taken the trouble of reproducing Jordan’s own words here, because it is essential that we understand the precise nature of the claim he is making.[2]

To begin, we define a projection operator E, the the expectation value of which, $\langle E\rangle$, is “the probability of a particular result for a measurement on one subsystem.” To picture what this operator does, suppose that we are measuring some quantity which has a non-degenerate spectrum of eigenvalues, and let ψ be the state of the system. Then in this case

$$\begin{aligned}\langle E\rangle &= \text{probability of a particular result } \beta \\ &= |\langle\beta|\psi\rangle|^2 \\ &= \langle\psi|\beta\rangle\langle\beta|\psi\rangle \end{aligned} \tag{3.1}$$

But

$$\langle E\rangle = \langle\psi|E|\psi\rangle. \tag{3.2}$$

Therefore in this case

$$E = |\beta\rangle\langle\beta|. \tag{3.3}$$

That is, E is simply the projector onto the eigensubspace of the operator we are measuring.

Now suppose a measurement of some other observable is made on the other subsystem, with a complete, discrete and nondegenerate spectrum of possible measurement results α_k. To each possible result there corresponds a projector $I_k = |\alpha_k\rangle\langle\alpha_k|$.

[2] All references in this section are from Jordan (1983: 264).

Because the $|\alpha_k\rangle$ define an eigenbasis of the second subsystem, we have

$$\sum_k I_k = \mathbb{1}. \qquad \text{(Closure relation)} \tag{3.4}$$

Again, we can think of the expectation values $\langle I_k \rangle$ as the probabilities of getting the corresponding eigenvalues α_k.

Now we seek the joint probability of finding measurement result β in one subsystem given a measurement result α in the other subsystem. Here is the key step: *"Since the measurements are on different subsystems, each I_k commutes with E."* (The emphasis is mine.) This means that the expression for joint probability will be just $\langle I_k E\rangle$. Then the probability of getting result β "regardless of what result is found in the other" subsystem is the sum of the joint probabilities for all possibilities; i.e.,

$$\begin{aligned} \sum_k \langle I_k E\rangle &= \left\langle \sum_k I_k E \right\rangle \\ &= \langle E\rangle. \end{aligned} \tag{3.5}$$

That is, the probability of getting a given measurement result in one subsystem is entirely independent of what, if anything, is done to the other subsystem. (This will not in general be true, of course, for correlations *between* measurement results.) But this is just the condition for no-signalling that we identified earlier. Hence there is no signalling between subsystems of a composite system *so long as* the subsystems are separate or separable in the sense Jordan indicates here; that is, in the sense that each subsystem can indeed be thought of as a distinct system.

What does "distinct" really mean? It means, just distinct enough for the projectors Jordan defines to commute. What is the physical meaning of this condition, however? Fine's representative comment is useful (from 1989):

> [QM]. . . does not allow the joint distribution of observables to be defined as a function of the state of the system. . . unless that state is one for which the observables are compatible. [p 179]

In other words, the joint probability $\langle I_k E\rangle$ that Jordan calculates could not even be defined unless the measurement on the subsystems were compatible. Compatible observables are those which can be measured simultaneously because measurement of one does not interfere with or interact dynamically with measurement of the other; the mathematical fact that they commute expresses the physical fact that the order in which they are applied to the system makes no difference to the outcome. (Perhaps Jordan thought it was too obvious to spell this all out.) In other words, Jordan's proof incorporates some very strong assumptions about the independence of the two subsystems; is it surprising that we get a no-signalling result? We will see that essentially the same assumptions are used in all of these arguments.

(Note that the expression for joint probability can be expanded as follows:

$$\begin{aligned} \langle I_k E\rangle &= \langle \psi | I_k E | \psi\rangle \\ &= \langle \psi|\alpha_k\rangle\langle\alpha_k|\beta\rangle\langle\beta|\psi\rangle. \end{aligned} \tag{3.6}$$

An expression of this form will appear in other proofs we shall consider.)

3.2 Some Properties of Density Operators

In order to understand the next two proofs we consider, it will be useful to review certain basic properties of the density operator. For comprehensive review, see Cohen-Tannoudji *et al.* (1977) (Vol. 1), Ballentine (1990b), or many other sources.

The *density operator* (also sometimes called the *statistical* or *state* operator) of a system in a pure state $|\psi\rangle$ is defined as

$$\rho = |\psi\rangle\langle\psi|. \tag{3.7}$$

If the system is in one of several possible states $|\psi_k\rangle$ with a probability p_k of being in the k-th state, then the density operator of the so-called mixed state is

$$\rho = \sum_k p_k |\psi_k\rangle\langle\psi_k| = \sum_k p_k \rho_k. \tag{3.8}$$

Some authors prefer to take the density operator as the basic object of study, rather than the state vector; see, for instance, Ballentine (1990b).

The expectation value of an operator A is calculated according to the useful formula

$$\langle A \rangle = \mathrm{Tr}\rho A, \tag{3.9}$$

which holds for both pure and mixed states (so long as ρ is appropriately normalized; i.e., $\mathrm{Tr}\rho = 1$.)

The reduced (or partial) density operator is essential for studying composite systems. (I rely here on the very clear discussion in Ballentine 1990b.) Let ρ_{LR} be the density operator of a composite system $S = L + R$. Then we define the reduced density operator for the left system as

$$\begin{aligned} \rho_L &= \mathrm{Tr}^{(R)} \rho_{LR}. \\ &= \sum_j \langle r_j | \rho_{LR} | r_j \rangle. \end{aligned} \tag{3.10}$$

The significance of the reduced density operator is that it can be used to compute probabilities and expectation values of observables in a subsystem of a composite system, *provided that* those observables are restricted in their operation to the subsystem. To see how this works, suppose the state of the composite system is spanned by vectors $|l_i r_j\rangle$, and let O_L be an operator which acts only on the left system. Its action on the total system will be given by an operator $O_{LR} = O_L \otimes \mathbb{1}_R$, with expectation value

$$\begin{aligned}
\langle O_{LR}\rangle &= \mathrm{Tr}\rho_{LR}O_{LR} \\
&= \sum_{i,j,i',j'} \langle l_i r_j|\rho_{LR}|l_{i'}r_{j'}\rangle\langle l_{i'}r_{j'}|(O_L\otimes\mathbb{1}_R)|l_i r_j\rangle \\
&= \sum_{i,j,i',j'} \langle l_i r_j|\rho_{LR}|l_{i'}r_{j'}\rangle\langle l_{i'}|O_L|l_i\rangle\delta_{jj'} \\
&= \sum_{i,i'}\sum_j \langle l_i r_j|\rho_{LR}|l_{i'}r_j\rangle\langle l_{i'}|O_L|l_i\rangle \\
&= \sum_{i,i'}\langle l_i|\Big(\sum_j\langle r_j|\rho_{LR}|r_j\rangle\Big)|l_{i'}\rangle\langle l_{i'}|O_L|l_i\rangle \\
&= \mathrm{Tr}^{(L)}\rho_L O_L. \qquad (3.11)
\end{aligned}$$

3.3 Ghirardi, Rimini and Weber (GRW)

The widely cited argument of Ghirardi, Rimini, and Weber (1980) uses a somewhat different approach than Jordan.

We consider two subsystems L and R of a composite system $S = L + R$ which we assume are widely separated in space and do not interact (although we allow the possibility that they may have interacted in the past). To say that they do not interact means that they do not interact dynamically; that is, they do not exchange mass-energy. This does not mean that the amplitudes associated with them may not *interfere* (which could be the case if they have interacted in the past). We also assume that in close proximity to L is a measuring apparatus A, and in close proximity to R is another apparatus B. (See the figure.)

Now, we write the density operator for the whole system before any measurement is made in this form:

$$\rho^0{}_{\text{total}} = \rho^0{}_{LR}\otimes\rho^0{}_A\otimes\rho^0{}_B, \qquad (3.12)$$

where $\rho^0{}_A$ is the initial density operator for the left apparatus, $\rho^0{}_B$ is the density operator for the right apparatus, and $\rho^0{}_{LR}$ is the density operator for the two subsytems. It is not written in a factorized form, to express the fact that these systems did interact at one time and may therefore continue to be correlated.

A point that is especially interesting about the GRW proof is that they do not assume any state reduction; their object is to show that QM "governs both the evolutions of the system and of measuring apparatuses, as well as their interactions" in such a way that the theory displays no "internal inconsistency (superluminal transmission of signals)" (p. 294).[3] To do this they "assume that the measuring process obeys Schrödinger evolution" (*loc. cit.*) and represent the effect of the measurement operations on the density operators by unitary operators; reduction involves a nonunitary process.[4]

[3] They don't explain why this would be an internal inconsistency; presumably they are referring to the causal paradoxes associated with superluminal signalling.

[4] Some authors argue that reduction is a blemish on the elegant scheme of QM; see, for instance,

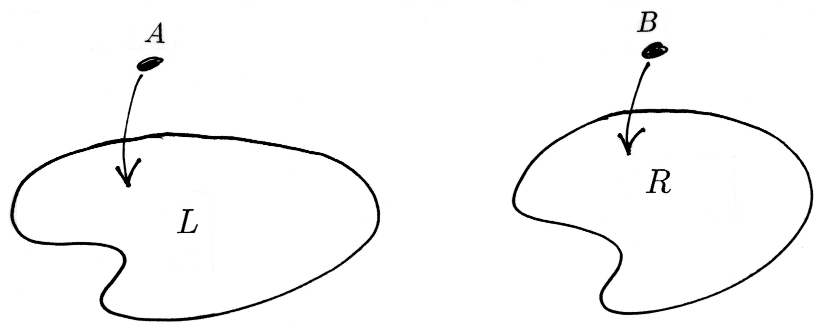

Figure 3.1: Systems and Apparatuses for the GRW Argument

Now, what we do is compare measurement results on one subsystem (say L) depending upon whether or not a measurement was made on R. Suppose the measurement by A on L is represented by a unitary operator U_{LA}, and the measurement by B on R is represented by a unitary operator U_{RB}.

After A interacts with L, but if B did not measure R, we get

$$\rho^A{}_{\text{total}} = U_{LA}\rho^0{}_{LR} \otimes \rho^0{}_A \otimes \rho^0{}_B U^\dagger{}_{LA}. \tag{3.13}$$

However, if the A measurement took place after the B measurement, we get

$$\rho^{BA}{}_{\text{total}} = U_{LA}U_{RB}\rho^0{}_{LR} \otimes \rho^0{}_A \otimes \rho^0{}_B U^\dagger{}_{RB}U^\dagger{}_{LA}. \tag{3.14}$$

Let O_A be an observable associated with A alone. We calculate its expectation value given measurement A followed measurement B, and show that it equals the expectation value given that only a measurement by A was performed.

$$\begin{aligned} \langle O_A\rangle^{BA} &= \mathrm{Tr} O_A \rho^{BA}{}_{\text{total}} \\ &= \mathrm{Tr} U^\dagger{}_{RB}U^\dagger{}_{LA}O_A U_{LA}U_{RB}\rho^0{}_{LR} \otimes \rho^0{}_A \otimes \rho^0{}_B. \end{aligned} \tag{3.15}$$

(The last step is done using the cyclic property of the trace: i.e., TrABC = TrCAB.) Now GRW say that since U_{LB} "commutes *by assumption*" with U_{LA} and O_A (even though they never actually stated any such assumption) we get

$$\begin{aligned} \langle O_A\rangle^{BA} &= \mathrm{Tr} O_A U_{LA}\rho^0{}_{LR} \otimes \rho^0{}_A \otimes \rho^0{}_B U^\dagger{}_{LA} \\ &= \mathrm{Tr} O_A \rho^A{}_{\text{total}} \\ &= \langle O_A\rangle_A. \end{aligned} \tag{3.16}$$

Again, we get statistical independence of the left measurement result upon the right measurement procedure, using commutation of measurement operators at a crucial juncture in the derivation.

Ballentine (1990a) or DeWitt & Graham (1973). Others argue that blemish or not, that's just the way things are; see Penrose (1989).

3.4 A General Density Operator Proof

A very compact proof is available using density operator notation. I give a version due essentially to Ballentine & Jarrett (1987).

Let x_L, x_R be possible measurement results (eigenvalues of some operator) on the left and right systems (particles) respectively. First, we write the probability that we will find particular values x_L and x_R for a particular pair of particles:

$$\begin{aligned} P(x_L, x_R) &= \mathrm{Tr}|x_L x_R\rangle\langle x_L x_R|\rho_{LR} \\ &= \langle x_L x_R|\rho_{LR}|x_L x_R\rangle. \end{aligned} \tag{3.17}$$

The reduced density operator on the left particle is

$$\begin{aligned} \rho_L &= \mathrm{Tr}^{(R)}\rho_{LR} \\ &= \sum_{x_R}\langle x_R|\rho_{LR}|x_R\rangle. \end{aligned} \tag{3.18}$$

We use this to compute the probability of getting a particular result x_L on the left particle:

$$\begin{aligned} P(x_L) &= \mathrm{Tr}^{(L)}\left[|x_L\rangle\langle x_L|\rho_L\right] \\ &= \langle x_L|\Big(\sum_{x_R}\langle x_R|\rho_{LR}|x_R\rangle\Big)|x_L\rangle \\ &= \sum_{x_R}\langle x_L x_R|\rho_{LR}|x_L x_R\rangle \\ &= \sum_{x_R} P(x_L, x_R). \end{aligned} \tag{3.19}$$

But this is just the condition for Parameter Independence that we were seeking.

Ghirardi (1988) provides a density operator proof as well. I will skip the details of his algebra, which is slightly different than above. I will just note his crucial assumption (p. 95), which is:

> For an ensemble of composite systems $S = S_1 + S_2$, associated with the state $\rho(1,2)$, all physical predictions concerning the ensemble of the subsystems S_2 are derived through the statistical operator $\rho^{(2)} = \mathrm{Tr}^{(1)}\rho(1,2)$, where $\mathrm{Tr}^{(1)}$ denotes partial tracing on the variables of system 1.

But, as we have seen, the partial trace can be used only when the observables concerned are confined to the subsystem; again, we have an assumption that measurement is a fully local process.

3.5 Shimony's Time Evolution Operator Proof

Shimony (1984) provides an interesting variant, which is probably one of the most thorough and detailed NRQM proofs. I'll omit most of the steps; suffice it here to note that he assumes localizability at two points in the argument. First, he assumes that the

Hamiltonian for the total system (left and right systems and left and right apparatuses) factorizes:

$$H_{\text{tot}} = H_L \otimes \mathbb{1}_R + H_{RB} \otimes \mathbb{1}_L, \tag{3.20}$$

where H_{RB} is the Hamiltonian for the right system and apparatus combined. This implies that the time evolution operator for the system factorizes:

$$\begin{aligned} U(t) &= e^{iH_{\text{tot}}t} \\ &= e^{iH_L t} \otimes e^{iH_{RB}t}. \end{aligned} \tag{3.21}$$

He then sets out to calculate the expectation value of an operator G measured in the left system, and he assumes that this operator acts on the left system alone. Using

$$\psi(t) = U(t - t_0)|\psi(t_0)\rangle, \tag{3.22}$$

and

$$\langle G\rangle = \langle\psi(t)|\mathbb{1}_R \otimes G|\psi(t)\rangle \tag{3.23}$$

he shows that the expectation value $\langle G\rangle$ depends only upon the initial state of the left system and its Hamiltonian H_L.

Hence we see the basis of Shimony's comment (1986: 191), that "quantum mechanical predictions concerning ensembles of pairs of particles do not violate Parameter Independence, provided that nonlocality is not explicitly built into the interaction Hamiltonian of the particle pair."

3.6 Summary of the Strategies Available

As Eberhard & Ross (1989) point out, what these arguments all comes down to is assuming the dynamic independence of the two systems. As far as we know, there is phase entanglement and therefore nonclassical correlation between systems only if dynamical interactions have occurred or continue to occur between those systems. However, in the measurement theory of NRQM, the following things are assumed:

1. The two subsystems can be sufficiently separated in space so that after some time they do not interact dynamically.

2. There is no dynamical interaction between the left measuring apparatus and either the right measuring apparatus or the right system.

Hence the left apparatus will be uncorrelated with the right system and the right apparatus, and the right apparatus will be uncorrelated with the left system and left apparatus. The two systems L and R may still be correlated, of course, because they did interact at one time.

Mathematically, we can express this separability, or lack of correlation, between the apparatuses and subsystems in several ways.

- Any dynamic interaction—i.e., measurement—of the left apparatus upon the left system, or of the right apparatus upon the right system, has no effect upon the other system; this is expressed by saying that if an operator O_L acts upon or pertains to the left system, its effect on the total system is an operator

$$O_{LR} = O_L \otimes \mathbb{1}_R. \tag{3.24}$$

(Similarly for the right measurement, of course.) This localizability condition is sufficient to give the no-signalling property, since the reduced density operator can then be used to compute probabilities and expectation values in the left system associated with O_L, as in Sect. 3.3.1 above. It also gives us the commutation property used by Jordan and GRW, since any two such operators commute in their action on the total system. Here is a simple proof. Let $|lr\rangle = |l\rangle \otimes |r\rangle$ be a typical basis vector of the total system. Then

$$\begin{aligned}(O_L \otimes \mathbb{1}_R)(O_R \otimes \mathbb{1}_L)|lr\rangle &= (O_L \otimes \mathbb{1}_R)\left(|l\rangle \otimes O_R|r\rangle\right) \\ &= O_L|l\rangle \otimes O_R|r\rangle \\ &= (\mathbb{1}_L \otimes O_R)\left(O_L|l\rangle \otimes |r\rangle\right) \\ &= (O_R \otimes \mathbb{1}_L)(O_L \otimes \mathbb{1}_R)|lr\rangle \end{aligned} \tag{3.25}$$

This is likely what Jordan had in mind.

It is quite important to note that the converse of the above is not true; that is, the commutativity of operators does not imply their localizability to a subspace of the state space. This issue becomes central when we discuss the field theoretic proofs.

- The density operator for the total system will factorize, as in the GRW proof above.

- The Hamiltonian will factorize, as noted in Shimony's proof above. This is a special case, of course, of the first point above. In fact, all of these proofs follow from the localizability of the operators pertaining to the subsystems; the details of the proof depend mainly upon which operator it is—be it the Hamiltonian, the time evolution operator, the density operator, etc.—that one wishes to examine.

3.7 Why NRQM Proofs Don't Do the Job

It may seem perverse to call these arguments into question, since they just seem to follow from very straightforward and time-tested applications of basic NRQM and its associated measurement theory. However, I think we can make a case that they do not accomplish what they have been touted as having accomplished—namely, provide a foundation for peaceful coexistence. They *do* describe what is undoubtedly true for a wide range of familiar quantum systems, but we will see, I think, that this is not quite enough.

Our problem is whether quantum theory calls into question the relativistic picture of causation. In other words, can measurements—which are just a subclass of dynamical interactions between systems of different scales—be nonlocal or have nonlocal influences? We have already seen a good indication of the fact that present-day quantum measurement theory essentially defines this problem out of existence by assuming that measurement is always a localized process. But there is no real justification for this expedient procedure; as Stapp told us in the remark quoted in Chapter 1, we are nowhere close to understanding the correct way to apply quantum concepts at the mesoscopic and macroscopic level, the level at which quantal systems interact with those human-scale systems we conveniently call measuring devices.

All the proofs of the NST within non-relativistic QM incorporate very strong assumptions about the dynamic separability of subsystems and measuring apparatus. Reasonable and plausible these assumptions may be, but they are there. Therefore, the proofs that we have examined so far all seem to simply say this: *if the measurement process is localized and if the two subsystems are dynamically localized, then an easy calculation shows that there is no violation of Parameter Independence and thus no controllable signalling* (controllable signalling being signalling that is dependent upon the choice of measurement process.) This certainly helps to clarify what is implied by our notions about localizability. But the PPC is significant only as a categorical proposition, not as a conditional proposition. Yes, it is surely true that a truly nonlocal measurement would be a very odd thing in the light of our present experience; the truly significant question, though, is whether it is possible, not whether there might be a violation of relativistic causation *in spite of* a presumption of the complete *localizability* of all measurements. *That* problem is very much a "straw man".

Another way of describing what we do in these proofs is assume that the state of the particle is a direct product of its position states of the particle and its spin states. This is the same thing as saying that we presume that there is no interference between its spin and position states. But clearly, QM offers no general warrant for this; it is an arbitrary restriction on the possible states of the particle, justifiable only by its general reasonableness in the light of our ordinary experience. It seems just as reasonable as the locality assumptions which go into the derivation of the Bell Inequality...

Recall again Shimony's statement that there is no signalling,

> ...provided that nonlocality is not explicitly built into the interaction Hamiltonian of the particle pair.

But why *shouldn't* nonlocality be built into the description of the system, or at least not explicitly be built *out?* The formalism certainly allows it. The whole point of the investigation is to clarify when we are entitled to treat quantum systems as local. There is nothing in the formalism of QM which says that any system can always be conveniently decomposed into spatially separate, separable or factorizable subsystems, or which would prohibit interference between position and other states, between a measuring apparatus and other subcomponents of a system, or that would prohibit a nonlocal Hamiltonian.

Fine (1989: 183) says,

> . . . a principle denying *any* influence between happenings in different wings . . . is actually built into the quantum theory, according to which there *is* no influence between the two wings of the experiment, that is, no physical interaction of any sort that is represented by terms in the Hamiltonian of the composite system. . .

But this seems to be just false. It is true that quantum measurement theory generally takes measurement operators to be localizable, but it is highly arguable that this provision is "built into" quantum theory; quantum measurement theory is a set of ingenious but (to some degree at least) provisional and *ad hoc* side conditions imposed upon the formalism of QM itself in order to give us some means of navigating the murky ground between the microscopic and the mesoscopic. By saying this I certainly do not mean to deny that there were excellent reasons why measurement theory was structured as it was by its creators; I am simply insisting that it is provisional and supplementary, and can be expected to be revised as our understanding increases.

Can we see anything else in these arguments, therefore, than a tenacious committment to Einstein's *Trennungsprinzip*, a principle which must certainly be left open to question in any discussion of the nature and limitations of nonlocality?

Well, yes, we can. First, it just happens to be our experience that all the typical experimental apparatuses that we are familiar with—magnets, polarizers, etc.—can be arranged in such a way as to have zero or vanishingly small dynamical interaction with each other. Second, it happens to be our experience that particles that have once interacted can be made to have vanishingly small interaction if allowed to separate sufficiently. And third, we have a very good theory, the theory of relativity, which tells us that if the particles get far enough apart, they can be dynamically independent for a time at least as long as the time required for light to travel between them.

Recall Bohr's statement in his reply to EPR:[5] "There is in a case like [EPR] no question of a mechanical disturbance of the system under investigation during the last critical stage of the measurement procedure." Bohr did not see the necessity of justifying this statement, but if he had, he likely would have said something like the above. Wouldn't that be good enough?

Unfortunately, in this context, no. I do not object to the appeal to experience; it just helps to confirm our faith that ordinary physics is good For All Practical Purposes (FAPP as Bell (1990) puts it). But we are concerned with an issue of principle here: whether or not QM can provide a logically independent justification of the relativistic picture of causation. It absolutely will not do to *appeal* to relativity in order to defend it. Recall what Copi said: if we just concede that relativity has to be imposed on QM as a side condition, then there is no point in trying to prove a no-signalling result (except perhaps as an indication of the consistency of our formalism). If we think we can derive a no-signalling result from QM, then we had better not use some kind of assumptions about no-signalling or no superluminal transmission at some point in the proof.

[5] In Wheeler & Zurek (eds.) (1983).

I do certainly believe that the NST does apply to all the set-piece entangled systems we have studied so far—the singlet state, for instance. I do not think that there is any ingenious arrangement of mirrors, lasers, magnets, or other hardware that we are presently familiar with that can defeat the simple proofs outlined in this chapter. If there are superluminal interactions, they will come about in a different way, and they will not be energetically free. In spite of this, the attempts to prove the NST *in all generality*—and not just for the localized garden-variety physics with which we are familiar—seem to be question-begging. Therefore, although I think the NST is almost certainly true (at least, taken in the sense that one cannot get "free" signalling by merely twiddling a polarizer) I am not convinced that we entirely know the reason why it is true, nor that we have a convincing proof of the general proposition.

Incidentally, Arthur Fine (1989: 183–185) of all authors, comes closest to acknowledging what I am worried about here; he is the only author I have seen who comes close to acknowledging that the dependence of no-signalling arguments upon a localizability premise smacks of question-begging. His comments are worth quoting (Fine 1989: 185):

> [A possible objection is that] The argument contains an undischarged assumption: namely, that performing a measurement on one wing does not influence what measurement is performed in the other wing.

However, he does not see the obvious and very serious implication of this objection, namely, that the no-signalling proofs may be entirely circular. Instead, he is diverted by the worry that if we think of mysterious nonlocal influences extending between distant measuring apparatuses the apparent ability of the experimenter to choose which measurement to make might be an illusion:

> To take this objection seriously would be to open up the possibility of what John Bell (Davies and Brown 1986, p. 47) calls "superdeterminism". It is a skeptical hypothesis that, once opened, could not easily be laid to rest. But like other skeptical hypotheses, it would require a lot of work, I believe, to get it going. Briefly, in the absence of a theory of "influences" between measurements performed, which shows how to integrate them with ordinary experimental practice, and in the absence of specific reasons to entertain suspicions about the presence of such influences, I think we can pass them by. Merely skeptical doubts ought not to stand in the way of judgements based on otherwise sensible arguments.

Whether or not we are inclined to lose sleep over the possibility of "superdeterminism" (and I am not), we should note that Fine's remarks above—which are roughly of the "that's a reasonable assumption anyway" variety—could be directed against other "skeptical hypotheses" like my own. I don't quite know what to say about this sort of response, apart from reiterating my basic and very simple claim, which is that a circular argument does not establish the thing that it purports to establish. This may seem like barren skepticism, but it is also sound informal logic.

Here is another way of looking at it. The crucial assumption in these proofs is the dynamical independence of the two systems—the lack of interaction (i.e., energy exchange) between them, which is expressed mathematically by the separability of the Hamiltonian for the total system. In other words, these proofs are saying one thing that is certainly very reasonable: namely, there is no thermodynamically *free* signalling. However, the strongest argument in favour of the dynamic independence of the two wings of the experiment is the one argument that *we do not have access to in this discussion*—namely, the relativistic prohibition on interactions outside the light cone.

It is perfectly true that we *know of* no interactions (gravitational, electromagnetic, nuclear) which do not fall off with distance.[6] We *know of* no superluminal *energy-bearing* entities or processes (tachyons, if you like) which could maintain dynamic interaction between the systems over arbitrary spacelike separation.

But all this means is that there is no superluminal signalling for any normal, reasonable systems with which we might presently be familiar; this is not a surprise. Obviously, though, we cannot use this generalization as a defence of relativity. That would be a proof that there is no violation of relativity, based upon the assumption that there is no violation of relativity. Yes, there is certainly no signalling if the total Hamiltonian of the system and apparatuses is local; and that is true if and only if there is no exchange of energy or information between the remote parts of the system. *But wasn't that what we were supposed to prove??*

Perhaps we could put it this way: I am saying that you (an author of a non-relativistic no-signalling proof) are merely saying that if there is no signalling, then there is no signalling. And you could very well reply in your defence, look, all *you* are saying is that if there is signalling then (because my separability condition will be violated) there is signalling.

The real point is that within non-relativistic QM we cannot get an answer to the signalling question without making some additional assumptions, *no matter how reasonable*, above and beyond what the formalism tells us. Therefore it does not seem to be true that the NST is "a simple consequence of quantum rules"[7] *alone*. What we really need is a general analysis of processes outside the light cone. This leads us to relativistic quantum mechanics and quantum field theory, which provide a broad enough framework that we can begin to get a grip on the problem of analysing the behavior of these nonlocal processes in spacetime.

[6] Except for the interaction between quarks. Are we justified in thinking that to be of no importance in this discussion?

[7] Ghirardi 1988: 95.

Chapter 4

Proofs from Quantum Field Theory

> ...relativistic quantum field theory, ...*is constructed to ensure* that its predictions do not depend either on the frame of reference or upon the order in which one imagines performing measurements on spacelike separated regions.
>
> — H. P. STAPP (1988: 376)

One of the most widely cited proofs of the NST was given by Eberhard (1978); this paper was one of the first, if not the first, to present a fully general no-signalling result that was directed specifically to the discussion of quantum nonlocality. (Everett (1973: 82–83) gives a no-signalling proof couched in the special terms of his Many-Worlds Theory.) Eberhard's proof, like some others we have considered, depends upon commutation of measurement operators. The commutation is, however, justified in a significantly different way, namely by appeal to a principle from quantum field theory (QFT) usually known as microcausality or local commutativity. Eberhard's proof deserves consideration in some detail because the problem of justifying the use of microcausality leads us right to the heart of the question we are concerned with, the relation between relativity and the quantum.

The principle of microcausality can be stated accurately enough for now as follows:

> MC: Observables acting at or defined at spacetime points with a spacelike separation are always compatible, and so their associated operators always commute.

In this chapter we will see how this rule guarantees a no-signalling result; in the next chapter we will discuss the justification and history of MC.

Other proofs of the NST based upon microcausality, essentially the same as Eberhard's, have been given by d'Espagnat (1984) and Sánchez-Gómez (1988); Eberhard and Ross (1989) review several similar strategies within both QFT and NRQM for deriving a no-signalling result. I will give two versions of Eberhard's proof here: an elementary proof, and a general version using density operators.

4.1 The Eberhard Proof: An Elementary Version

Eberhard's (1978) proof is given in terms of density operators; this gives his derivation full generality, since the theorems of density operator theory he uses are believed to be true for either pure or mixed states, with degenerate or non-degenerate eigenvalues with either a continuous or discrete spectrum. I will first give an elementary version of his proof for the restricted case of systems in pure states with discrete, non-degenerate eigenvalues; this allows us to replace his density operators with projectors and makes the structure of the proof entirely transparent.

We consider some spatially-extended system described by a pure (normalized) state $|\Psi\rangle$. Note that normally we would tend to think of the base states $|a\rangle$ of $|\Psi\rangle$ as being in turn product states of the form

$$|a\rangle = |xy\rangle = |x\rangle \otimes |y\rangle \tag{4.1}$$

where $|x\rangle$ is some eigenstate of the part of the system acted upon by A and $|y\rangle$ is the corresponding eigenstate for the remote part of the system. (For instance, we believe that the singlet state is a superposition of states of the form $|+-\rangle$ and $|-+\rangle$.) But it is not absolutely clear that we are compelled to write $|a\rangle$ as in (4.1), even though it is hard to see what it would mean from an operational point of view for a localized measurement to drive a system into a global state that was not a product state. However, the lesson of twentieth century physics is this: all physical systems are nonlocal in the first instance (i.e., their properties are functions of the *whole* system); it is only in certain circumstances that the system may be separable with respect to space or some other observable considered as a parameter. Therefore we may actually be assuming more than we are entitled to, if we assume that the base states of $|\Psi\rangle$ are necessarily in product form (i.e., separable). The global state $|\Psi\rangle$ could be some superposition, factorizable or otherwise, of product states, such as the singlet state, but it is at least thinkable that it could also be a state which is not expressible at all as a linear combination of products of eigenstates of localized systems; the important point for the argument here is that it is spatially extended. A great merit of Eberhard's formalism is that it permits this greater degree of generality, and, as we will see, thereby exposes very clearly the role that MC plays in the proof.

Now, assume that there are two localized detectors A and B, positioned so as to interact with spacelike separate parts of the system. A detector is any device such as a polarizer or Stern-Gerlach apparatus which can interact with some system of interest, together with some sort of recording device. At its location (which we shall call one branch of the system), A measures some observable with possible values a, while at the other branch B measures an observable (possibly different) with possible values b. Associated with each a and b there are eigenstates $|a\rangle$ and $|b\rangle$ of the extended system; we will assume that these eigenkets are normalized and define discrete, non-degenerate bases of the extended system. The nature of the "association" could, of course, be problematic.

Now, suppose A makes its measurement first. (What "first" means depends, of course, upon the reference frame, but it will turn out that because the results are entirely symmetrical between A and B this does not matter.) The state vector $|\Psi\rangle$ for the composite system will collapse[1] to a reduced state $|\Psi'\rangle$, and if we want to calculate the probabilities of the possible results of the measurements made by B *afterwards* we have to use $|\Psi'\rangle$. As usual, our object is to calculate the probability that B's measurement yields a particular value $b = \beta$ for *any* value of a previously found by A, and show that this is equal to the probability that $b = \beta$ in the case that *no* measurement was made by A. If we can do this we will have shown that there is no net statistical influence on B's results by A's measurements.

Assume that A gets some particular result α. The probability of this happening is

$$\mathrm{Prob}(a = \alpha) = \left|\langle\alpha|\Psi\rangle\right|^2. \tag{4.2}$$

By the projection postulate, the reduced state resulting from this measurement is

$$|\Psi'\rangle = |\alpha\rangle. \tag{4.3}$$

Then

$$\begin{aligned}\mathrm{Prob}\ (b = \beta\ \ \textit{after}\ A\text{'s measurement}) &= \left|\langle\beta|\Psi\prime\rangle\right|^2 \\ &= \left|\langle\beta|\alpha\rangle\right|^2.\end{aligned} \tag{4.4}$$

Note again that $|\alpha\rangle$ and $|\beta\rangle$ are *global* eigenstates, that is, eigenstates of the total apparatus, not of some localized part of the system (such as a particle) in the vicinity of A or B. This expresses the awkward fact of nonlocality, namely that a localized measurement can drive a spatially extended system into a definite global state.

Now we calculate the joint probability that A gets α and then B gets β:

$$\begin{aligned}\mathrm{Prob}\ (b = \beta\ \&\ a = \alpha) &= \left|\langle\alpha|\Psi\rangle\right|^2\left|\langle\beta|\alpha\rangle\right|^2 \\ &= \langle\Psi|\alpha\rangle\langle\alpha|\beta\rangle\langle\beta|\alpha\rangle\langle\alpha|\Psi\rangle.\end{aligned} \tag{4.5}$$

Now the crucial step: even though the projectors $|\alpha\rangle\langle\alpha|$ and $|\beta\rangle\langle\beta|$ are defined in terms of possibly global objects $|\alpha\rangle$ and $|\beta\rangle$, we assume that they represent the effect on the system of localized measurements taking place at A and B respectively. We

[1] Eberhard freely uses the concept of state reduction in his proof. Since Eberhard wrote, Ballentine (1990) and others have presented arguments against the projection postulate; we will see, however, that whether or not a concept of reduction is used is not the essential feature of the signalling question.

therefore assume that MC applies to them. Then we can interchange them in the above expression to get

$$\begin{aligned}\text{Prob}\ (b = \beta\ \&\ a = \alpha) &= \langle\Psi|\alpha\rangle\langle\alpha|\alpha\rangle\langle\alpha|\beta\rangle\langle\beta|\Psi\rangle \\ &= \langle\Psi|\alpha\rangle\langle\alpha|\beta\rangle\langle\beta|\Psi\rangle. \qquad (4.6)\end{aligned}$$

This joint probability—which is what enters into the nonclassical correlations that give rise to the violation of Bell's Inequality in such experiments—is influenced by the measurement at A. This expresses the fact that Outcome Independence is violated in such an arrangement.

(Note also that this expression for joint probability has the same form as the expression for the expectation value $\langle I_k E\rangle$ in Eq. (3.6) of Jordan's proof, Sect. 3.1.)

Now, at last, we can write the probability that a measurement by B gives β, given *any* possible outcome at A:

$$\begin{aligned}\text{Prob}\ (b = \beta\ \&\ a = \alpha_1 \vee\ \alpha_2 \vee \ldots) &= \sum_i \langle\Psi|\alpha_i\rangle\langle\alpha_i|\beta\rangle\langle\beta|\Psi\rangle \\ &= \langle\Psi|\mathbb{1}|\beta\rangle\langle\beta|\Psi\rangle \quad \text{(Closure relation)} \\ &= \langle\Psi|\beta\rangle\langle\beta|\Psi\rangle \\ &= \Big|\langle\beta|\Psi\rangle\Big|^2. \qquad (4.7)\end{aligned}$$

But, as expected, this is just the result that B would have obtained if *no* measurements were made by A.

The whole argument is just trivial manipulation of Dirac brackets; the only non-trivial step (assuming we accept the uncritical use of the projection postulate) is the interchange of the projectors $|\alpha\rangle\langle\alpha|$ and $|\beta\rangle\langle\beta|$. If we could not do this the calculation would have to stop at line (4.5), unless we had more specific information about the structure of $|\alpha\rangle$ and $|\beta\rangle$.

4.2 The Eberhard Proof: General Version

At the risk of being slightly repetitious, I will now give the general form of Eberhard's proof, essentially as given in Eberhard (1978). My notation will be very similar to Eberhard's, which is especially perspicuous.

As usual, we consider two systems S_L and S_R, which are widely separated in space but which may have interacted in the past. M_L and M_R are measurements made on the left and right systems respectively; L and R are what Eberhard calls "knob settings" (i.e. parameters) for the left and right measurements; the $\{l\}$ are possible left results (i.e., eigenvalues of whatever observable is measured) and the $\{r\}$ are possible right results.[2] We define $\Pr(l, L, r, R)$ as the joint probability that M_L yields l and M_R yields r when the parameters are L and R.

[2]The proof is a little more general than this terminology suggests, since L and R might even stand for the total states of the measuring devices. I thank J. R. Brown for this observation.

How do we define the no-signalling property? What we want to show is that the frequencies that (say) the left observer will see are statistically independent of whatever the right observer chooses to do. This means that the marginal probability of a given left result l—which will be the probability distribution of l summed over all possible r for a given right observable—should be independent of R. That is,

$$\sum_r \Pr(l, L, r, R) = F(l, L), \tag{4.8}$$

where $F(l, L)$ is some function of l and L *alone*. If this holds, then the left observer cannot tell what the right observer did because he cannot tell what results the right observer got, and no signalling is possible. A similar condition should, of course, hold for attempts by the left observer to signal to the right.

The reader may be puzzled by the fact that here and in other proofs we have covered we sum over *outcomes*, in spite of all the talk of *Parameter* Independence; I will return to this point shortly.

To calculate this joint probability, we define projectors $P(l, L)$ and $Q(r, R)$ associated with M_L and M_R respectively. These have the properties *Idempotence*; i.e.,

$$P^2(l, L) = P(l, L), \tag{4.9}$$
$$Q^2(r, R) = Q(r, R), \tag{4.10}$$

and *Closure*; i.e.,

$$\sum_l P(l, L) = \sum_r Q(r, R) = \mathbb{1}. \tag{4.11}$$

Now we use the following assumptions:

1. The probability of getting a result l is

$$\Pr(l, L, \rho) = \mathrm{Tr} P(l, L)\rho, \tag{4.12}$$

   where ρ is the density (statistical) operator for the system.

2. Suppose M_L is made "first". Then ρ reduces to ρ', where

$$\rho' = \frac{P(l, L)\rho P(l, L)}{\mathrm{Tr} P(l, L)\rho}. \tag{4.13}$$

   This is the Lüders Rule for state reduction.

3. The joint probability we seek is the product of the probability of getting l and the probability of getting r:

$$\Pr(l, L, r, R) = \Pr(l, L, \rho) \Pr(r, R, \rho'). \tag{4.14}$$

Only assumption 2 might be controversial, since a few authors (notably Ballentine 1990a,b) do not accept the notion of state reduction as legitimate or well-defined. Leaving this possible objection aside for now, we calculate:

$$\begin{aligned}\Pr(l,L,r,R) &= \mathrm{Tr}P(l,L)\rho\,\mathrm{Tr}Q(r,R)\rho' \\ &= \mathrm{Tr}P(l,L)\rho\,\mathrm{Tr}R(r,R)\frac{P(l,L)\rho P(l,L)}{\mathrm{Tr}P(l,L)\rho} \\ &= \mathrm{Tr}P(l,L)Q(r,R)P(l,L)\rho. \end{aligned} \tag{4.15}$$

In the last step we use the cyclic property of the trace. Now, as before, we employ the crucial "simplification" (as d'Espagnat calls it, 1984: 248) permitted by microcausality: we interchange Q and P in the last line above and get

$$\begin{aligned}\Pr(l,L,r,R) &= \mathrm{Tr}P^2(l,L)Q(r,R)\rho \\ &= \mathrm{Tr}P(l,L)Q(r,R)\rho. \end{aligned} \tag{4.16}$$

Eberhard (1978: 416) justifies this move by saying:

> In the formalism of quantum theory, locality is expressed by the requirement that the measurement operators outside of the light cone commute.

It is quite unclear what logical relation between locality and commutativity this statement means to indicate; it almost seems as if Eberhard is saying that MC expresses a prior *assumption* that locality holds. As we shall see, this is essentially the case.

Note also that Equation (4.16) states that the joint probability of a particular right result and a particular left result depends upon the choice of measurement process on the right (actually we can just as easily say that right depends upon left, since the expression is symmetrical in P and Q). This is the source of the non-classical correlations (such as $-\cos\theta_{\mathbf{ab}}$ in the singlet state) that violate the Bell Inequality; in other words, this demonstrates a violation of Outcome Independence.

Now, we go ahead and calculate the marginal probability we seek:

$$\begin{aligned}\sum_r \Pr(l,L,r,R) &= \sum_r \mathrm{Tr}P(l,L)Q(r,R)\rho \\ &= \mathrm{Tr}\,(\sum_r Q(r,R))P(l,L)\,\rho \\ &= \mathrm{Tr}P(l,L)\,\rho \\ &= \text{function of } l \text{ and } L \text{ only.} \end{aligned} \tag{4.17}$$

Similarly we can get

$$\sum_l \Pr(l,L,r,R) = \text{function of } r \text{ and } R \text{ only.} \tag{4.18}$$

Therefore, given MC, there is no statistical influence of the choice of observable on one side upon the probability of getting a particular result on the other side.

4.3 How to Signal

Now, let us suppose that there were some observables P, Q and Q' such that $Q \neq Q'$ and

$$\begin{aligned} [P(l,L),\, Q(r,R)] &\neq 0, \\ [P(l,L),\, Q'(r',R')] &\neq 0. \end{aligned} \tag{4.19}$$

even though S_L and S_R were spacelike separated.[3] (The symbols r' and R' of course refer to the eigenvalues and parameters associated with Q'.) Then the corresponding expressions for marginal probability will contain irreducible cross-terms:

$$\sum_r \Pr(l,L,r,R) = \operatorname{Tr} \sum_r \big(P(l,L)Q(r,R)P(l,L)\big)\rho \tag{4.20}$$

and

$$\sum_r \Pr(l,L,r',R') = \operatorname{Tr} \sum_r \big(P(l,L)Q'(r,R)P(l,L)\big)\rho. \tag{4.21}$$

Hence the phase entanglement would show up in the marginal probabilities; that is, the choice of observable on one side could affect the probability of getting particular results on the other side. Furthermore, in general we will have

$$\sum_r \Pr(l,L,r,R) \neq \sum_r \Pr(l,L,r',R'). \tag{4.22}$$

Therefore, to send a message from right to left, we could simply arrange that the left observer always measures the same observable, and then vary the choice of right observable in step with (say) Morse Code. (The simplest procedure would be to let Q' be identity; i.e., just let the particles pass through the detector without interaction.) The fact that this is, at least in principle, possible, given the failure of MC, indicates that given the background assumptions 1 to 3, MC is both a necessary and sufficient condition for no-signalling.

4.4 Is This Really Parameter Independence?

Eberhard's notation helps to emphasize a fact that the perceptive reader must have noticed by now, which is that although we have been talking all along of the no-signalling property as *Parameter* Independence, the parameters L and R are essentially irrelevant in the derivation. The summation in Equation (4.17) is over *outcomes* at the distant branch of the apparatus; the presumptive "knob settings" play no rôle at all *except* insofar as choice of parameter is equivalent to or conditions the choice of observable. (For instance, if I orient my detector in the z direction this is equivalent to choosing to measure S_z.) This irrelevance of parameters is expressed mathematically by the fact that the summation in the closure relation

$$\sum_r Q(r,R) = \mathbb{1} \tag{4.23}$$

[3]The expression $[A,B]$ denotes the commutator of A and B; i.e., $[A,B] = AB - BA$.

is over possible outcomes, not possible parameters. What counts is the distant experimenter's choice of *observable* to be measured, be it position, momentum, or spin in a certain direction. This determines the set of possible eigenvalues $\{r\}$ or $\{l\}$ which are actually used in the calculation.

4.5 The Question is Commutativity

We have already noted that Eberhard uses a reduction postulate, namely the Lüders rule. It is interesting to note that a NRQM proof of the NST given by Hughes (1989), which also uses the Lüders reduction scheme, is formally identical to Eberhard's. (Hughes, as does van Fraassen—see van Fraassen 1989—calls the No-Signalling property "Surface Locality".) Hughes, however, just takes it for granted that the operators are localizable (Hughes 1989: 251):

> In the case of a composite system, events associated with one system are always compatible with events associated with the other, since all pairs of operators $\mathbf{A} \otimes \mathbb{1}$ and $\mathbb{1} \otimes \mathbf{B}$ commute. [**A** and **B** are operators associated with the left and right systems respectively.]

Even though his algebra is formally the same as Eberhard's, I class his proof as falling within NRQM simply because he makes no special appeal to field-theoretic or relativistic principles; what counts is how one justifies the commutativity of measurements on the two subsystems. In the GRW proof (Sect. 3.3), which does not assume that reduction takes place, one could have just as easily justified commutativity of measurements by MC as by assuming the localizability of measurements; conversely, Parameter Independence would break down in the GRW picture just as easily as in the Eberhard picture if one could not commute the operators on the two subsystems. It is not really *state reduction* but phase entanglement that opens up the possibility of signalling, and it is the commutativity of spacelike separated observables that suppresses this entanglement. We now turn to the problem of understanding the basis for MC within QFT, and the interesting bearing this has on the argument for peaceful coexistence.

Chapter 5

Foundations and Limitations of QFT Proofs

> It seems to me that it is among the most sure-footed of quantum physicists, those who have it *in their bones*, that one finds the greatest impatience with the idea that the 'foundations of quantum mechanics' might need some attention. Knowing what is right by instinct, they can become a little impatient with nitpicking distinctions between theorems and assumptions.
>
> — J. S. BELL (1990: 33)

The major claim of this thesis is that the argument by Shimony and others in favour of peaceful coexistence as a categorical proposition is fallacious because the proofs of the NST that have so far been given in the literature are either *insufficiently general* or *circular*. I have already argued that the NRQM proofs of the NST are circular, or at least very questionably founded, since they depend upon such strong assumptions about the localizability of those interactions that we loosely term "measurements" that they seem to render themselves incapable of examining the very possibility they are are meant to eliminate. We shall see that the way the commutativity of spacelike separate operators is justified within QFT vitiates the argument for peaceful coexistence in an interestingly different way.

Before going further, it is useful to note that there is great variation in terminology in this subject. What I have been calling *microcausality* (MC) is usually taken to mean the following:

> Operators which represent measurements performed on space-like separate parts of a physical system always commute, regardless of whether or not they would commute if operating locally.

In many books on field theory, however, this is often referred to as *local commutativity* (see Bogolubov *et al.* 1975 or Streater & Wightman 1964, for instance); this name is meant to emphasize that this is a locality principle. In fact, it seems like *distant commutativity* might be a more descriptive name for this property, but I will stick to the more-or-less standard usage in order to avoid confusion. "Microcausality" is also sometimes used in a much more specialized sense which would take us far beyond the scope of this investigation to understand (see Bogolubov *et al.* 1975); in any case, the principle I have defined here, whatever we choose to call it, is the principle that is relevant to our discussion since it is what has been used in the no-signalling proofs given by Eberhard and others.

As we have already seen (see Sect. 3.6), there is an easy way to justify the commutativity of the operators associated with the measurements on the two branches of the system: simply assume that they must be local. Mathematically, this means that they act nontrivially only on the single subspace belonging to the particle that is being measured; commutativity of the operator in its global action follows automatically. In fact, MC is often spoken of as if it were simply an expression of the presumptive locality of measurements. For instance, Streater (1987: 141) says:

> *local commutativity*... means that observables commute when localized in spacelike separated regions.

If this were the whole story of MC the problem of its justification and the limitations of its use would be the same as the problem of the justification and limitations of the hypothesis of the locality of measurements; there would be nothing special about its use in a relativistic context. However, MC was actually introduced to the "formalism of quantum theory" (Eberhard's phrase) not primarily because of assumptions about the locality of observables (although these played a rôle) but as a consequence of attempts to make field theory consistent with relativity. It turns out that one must undertake quite an odyssey through the literature in order to get a clear picture of how this came about.

5.1 History and Basis of MC in QFT

In seeking the basis of MC within QFT, one might begin, for instance, with the generally very clear and useful book by Jauch & Rohrlich (1955). Unfortunately, when we seek an understanding of MC from these authors we are frustrated, for we find only this (Jauch & Rohrlich 1955: 11):

> ...it follows from general principles of quantum mechanics that two observables located at two points x_1, x_2 with a space-like separation... always commute. The physical content of this statement is that no measurable effect can be propagated faster than the speed of light. The measurement of quantities at space-like situated points must therefore be entirely independent of each other. This means in the quantum mechanical formalism that the observables associated with these quantities must commute.

It is most unclear what is supposed to follow from what in this statement; furthermore, what "general principles" of QM are the authors referring to? In fact, statements with this degree of logical clarity are not uncommon in this literature, and it turns out to be surprisingly difficult to assess the overall logical position of MC within QFT. This is complicated by the fact that QFT exists in a bewildering variety of formulations. In fact, it would be quite incorrect to say that there exists one coherent and axiomatized theory known as Quantum Field Theory. It is really a collection of rough-and-ready methods unified by some general guiding principles, which principles can and have been stated in a variety of ways.[1] Roughly speaking, though, it seems that from a survey of texts written over the last fifty years or so one can identify two (overlapping) phases in the history of MC. Earlier books, up to perhaps 1960 or so,[2] attempt to derive MC as a theorem from certain postulates about the form of commutators of field quantities.[3] After this time, there is a tendency to frankly accept MC as itself a postulate (although this approach in fact dates back at least as early as Pauli 1940). We shall say a little about both these views.

Mandl (1959) seems to be typical of the older way of thought. Field quantities in QFT are operators, defined in terms of the creation and destruction operators of the ordinary non-relativistic theory. These are assumed to obey the following commutation relation:

$$\left[a(\mathbf{k}_i),\, a^\dagger(\mathbf{k}_j)\right] = \delta(\mathbf{k}_i,\, \mathbf{k}_j), \tag{5.1}$$

(where the $\mathbf{k}$'s are momentum states). Mandl shows how to work from this to the form of the commutator for field quantities $\phi(x)$:

$$\left[\phi(x),\, \phi(x')\right] = i\Delta(x - x'). \tag{5.2}$$

The Δ-function is a relativistic generalization of the ordinary δ-function. It can be shown to be zero on any plane of simultaneity; therefore the field commutator always vanishes for any two points at a spacelike separation. This relation holds only for Bose particles; Mandl also shows how to derive an analogous expression for fermions, again starting from a postulated expression for creation and absorption operators involving a δ-function.

Mandl's comment is illuminating:

> The non-vanishing of the commutator [(5.2)] of two Hermitian operators means that the measurements of the corresponding observable fields at the spacetime points x and x' interfere with each other. But such interference cannot occur if $(x - x')$ is space-like; for this would require a signal to propagate between x and x' with a velocity greater than that of light, contrary

[1] See Chang (1990) for an up-to-date graduate-level introductory text on QFT; Bogolubov *et al.* (1975), Schweber (1961), Roman (1969) and Streater and Wightman (1964) discuss various ways in which the theory might be axiomatized; Streater (1989) gives a useful overview of the problem of axiomatizing QFT.

[2] E.g., Mandl (1959).

[3] In fact, this is still the practise in some more recent books, such as Lee (1981).

> to the theory of relativity. Thus the vanishing of the commutator [(10)] for space-like $(x - x')$ represents a most satisfactory link-up of quantum mechanics and the theory of relativity.[4]

We should note that it seems to be a very serious misunderstanding to suppose that the interference of quantum states is due to the propagation of any kind of *signal* at all (if this is indeed what Mandl is suggesting, for the sense of his term "require" is murky). But leaving this point aside (I shall return to it shortly), the procedure is clear: get the right commutators, and microcausality—and therefore the NST and peaceful coexistence—follows automatically.

There is one major objection to this line of thought: the δ- and Δ-functions used to define the field commutators above are short-hand for sequences of functions whose limits might not always have physical or operational significance. This could come about in two closely-related ways. Streater and Wightman (1964, p. 100–101) object that the field quantities will not be physically meaningful unless one adds to the theory a postulate that they should be written in "smeared" form—i.e., as an integral representation.[5] They point out that this would be an "additional strong assumption" which not only seems to lack sufficient general motivation but still fails to prevent some field quantities from being singular. Streater and Wightman conclude that "one is reluctant to accept canonical commutation relations as an indispensible requirement on a field theory".

One can add the more general objection that any function defined in terms of a limit could very likely fail to be physically significant in any case in which spacetime cannot be thought of as continuous. In fact, there certainly is some reason to believe that nature could be "nonlocal or granular in the small" (as Bjorken and Drell (1965) put it, p. 4), and perhaps not even merely in the small. (For instance, some recent speculations in cosmology assign a very low entropy to the early universe; this would only be meaningful if we could think of spacetime in the early universe as existing in a relatively small number of coherent states.)

Hence the defect of this older procedure is a lack of generality; while there is no doubt that it is essentially correct for a large part of the physics with which we are familiar, it does not produce a principle of MC which is broadly enough founded to be usable with confidence everywhere in field and particle theory.

From our point of view here, we can make this objection more pointed: to assume continuity of spacetime is in effect to assume that it has a light-cone structure,[6] and this is in turn to assume causal good behavior of the underlying manifold. This remark is quite tentative, because much mathematical analysis would be required to make it precise.) So, although no one puts it just this way, if we base the NST upon this older way of deriving MC, we *may* have essentially argued in a circle; we may have merely

[4] Mandl (1959), p. 29.

[5] Suppose some field quantity $\phi(\mathbf{x})$ is defined in terms of a δ-function. Like the δ-function, therefore, it is singular—it only has meaning when written under the integral sign as, e.g., $\int d\mathbf{x}\, f(\mathbf{x})\phi(\mathbf{x}) = \phi(f)$.

[6] My thanks to A. J. Stacey for this observation.

shown that relativistic causality is obeyed if it is obeyed. This is no doubt true, but hardly informative.

Apparently recognizing that MC considered as a theorem rested upon a somewhat precarious structure of assumptions, most quantum field theorists by about 1960 seem to have decided that it was far safer to simply *postulate* MC. The object of this postulation was simply to rule out of court from the outset any possible conflict between quantum mechanics and relativistic causality. Schweber—in the context of a discussion of the spin-statistics theorem—says this:

> The justification for the first postulate [MC] stems from the fact that in the quantum theory the lack of commutativity of two observable operators implies that these cannot be measured simultaneously with arbitrary accuracy. However, measurements at points which are separated by space-like distances can never perturb one another, since relativistic causality demands that signals (energy) cannot be propagated with a velocity greater than that of light. Hence, the commutator of observables with space-like connections must vanish. (1961: 223)

And further on in the same text, (Schweber 1961: 723):

> ...the microscopic causality principle...is the mathematical statement of the fact that no signal can be exchanged between two points separated by a space-like interval and therefore that measurements at such points cannot interfere.

Schweber is in fact echoing Pauli's seminal paper of 1940 on the spin-statistics theorem. In that paper, one of the founding works of modern quantum field theory, Pauli states:

> We shall...expressively postulate in the following *that all physical quantities at finite distances exterior to the light cone (for* $|x'_0 - x''_0| < |\mathbf{x}' - \mathbf{x}''|$*) are commutable.*...The justification for our postulate lies in the fact that measurements at two space points with a space-like distance can never disturb each other, since no signals can be transmitted with velocities greater than that of light. (1940, in Schwinger (ed.) 1958: 721.)

This is the clearest statement I have been able to find anywhere of the justification for the introduction of MC in field theory. Pauli's footnote to this statement is worth quoting as well:

> For the canonical quantization formalism this postulate is satisfied implicitly. But this postulate is much more general than the canonical formalism.

One way of seeing that Pauli's version of MC is more general than Mandl's is simply to note that it is not couched in terms of specific sorts of field operators $\phi_i(x)$, but *any* observables whatever; the motivation for this is that a difficulty for relativity arises if *any* spacelike separate observable fails to commute. In fact, as we have seen from considering Eberhard's proof, if we want to save textbook relativity we need to be able to apply MC even if we cannot make any assumptions about the localizability of the operators in question (for instance, whether or not they have the form $A_L \otimes \mathbb{1}_R$). There is nothing in the *formalism* that requires us to interpret the projectors $P(l, L)$ and $Q(r, R)$ or their eigenstates as localized, even though our interpretation does, of course, require that they have *some* locally observable effects among other effects that they might have. This is not such a bizarre conception as one might at first think: these operators could, for example, represent the creation of two or more particles at a spacelike separation; while they have a locally observable effect (the local appearance of one of the particles) the total process is nonlocal. If we want to apply MC to such operators we can hardly justify it by appeal to their locality. (I am not suggesting that Pauli or other theorists would necessarily have accepted this liberal view of the possible structure of operators, of course; it is simply that the possibility of this view seems to be implied by their procedure.) In fact, QFT is distinguished from ordinary quantum theory by the fact that it considers particle creation and destruction, and these are arguably inherently nonlocal processes; therefore QFT has to permit itself this degree of generality in formulating MC.

The upshot is that if we adopt the line of thought advocated by Pauli and later by Schweber and most other field theorists, our reasoning is the reverse of Mandl's: we *assume* that the relativistic view is correct and structure our rules about commutators accordingly. Note Pauli's use of the term *fact* to describe the statement that spacelike separate measurements cannot "disturb" each other; this is clearly taken as a *given*, not something to be established.

An important problem is to clarify what logical relation Pauli (and less clearly, Schweber and Mandl) is claiming to hold between the statement "no signal can be transmitted with velocities greater than that of light" (let us abbreviate this as NS) and the statement that "measurements at two space points with a space-like distance can never disturb each other" (let us just call this MC, since it is just a restatement of that rule). What is the meaning of Pauli's "since"? This can be read in two interestingly different ways.

On a very literal reading, Pauli and the others might be saying that any interference between two measurements would be due to exchange of signals or energy of some sort (faster than light particles, perhaps?); since this is outright prohibited by relativity, MC must hold. But I find it unlikely that Pauli, who was both a great master of quantum theory and a very logically precise thinker, could have held such a naïve view of the nature of interference between quantum mechanical states. To dream that interference could be due to exchange of particles[7] is certainly to put the classical cart

[7] And yet a tendency to use language that suggests something like this is very common even recently. For instance, Bell himself (Bell 1986: 9) says: "...the consequences of events at one place *propagate* to other places faster than light."

before the quantum horse, since particle states are defined in terms of superpositions of states.

What Pauli almost certainly had in mind was the following simple application of *Modus Tollens*:

$$NS, -MC \rightarrow -NS \vdash MC. \tag{5.3}$$

That is, given the *fact* (Pauli's term) of no-signalling, MC must hold because otherwise signalling would be possible. I don't think it is odd that Pauli did not trouble to state a proof of the conditional in this argument; he was famed for his impatience with spelling out what was (to him) the obvious.

The question now arises as to whether, given this historical background to the introduction of MC to QFT, we can really take Eberhard's proof as a sufficient defence of the PPC, as Shimony (1986) says we should.

5.2 The Logical Problem

If my analysis of Pauli's reasoning is correct, then perhaps we could say that what Eberhard and others have really done is not proven NS as a general property of quantum theory, but merely filled in the steps required to establish the conditional in (5.3); in other words, they have simply shown that Pauli's postulate, in conjunction with the other rules of QM, really accomplishes what it was meant to accomplish (i.e., rule out signalling). This certainly tells us that the QFT-based proofs of the NST are of more limited scope than people generally think, but does it mean that they are *circular?*

Recall that we defined an argument as circular if it has the general form

$$A, B \vdash A, \tag{5.4}$$

where B is the conjunction of a number of premisses we will call the *background conditions.* Eberhard's argument has the overall form

$$B, QFT, MC \vdash NS, \tag{5.5}$$

where by QFT we will mean the conjunction of the axioms and rules of inference of local quantum field theory except for MC itself, and where B are background premisses which could include elements of other physical theories which are taken as given. The question is, what statements properly belong to QFT and to B? This, of course, varies from author to author, from book to book, and in some cases (Schweber 1961, for instance) from chapter to chapter of a given book. In *Pauli's* version of QFT, MC is added to QM as an additional postulate simply in order that QM should be consistent with a prior acceptance of NS; therefore, obviously, if we think of ourselves as accepting what Pauli accepted then NS belongs to B and our overall argument is manifestly circular.

A defender of peaceful coexistence might well reply to these objections as follows: "I concede that as a matter of *historical* fact the creators of QFT openly accepted relativity as a given and carefully structured QFT so as to avoid conflict with it. But we do not have to endorse their motives in order to employ MC as a primitive postulate in an axiomatic local field theory. Our formal theory does not have to include NS as a *postulate* (even though we are not inclined to doubt it); then (by Eberhard's proof) our theory gives a logically independent validation of NS."

This procedure would remove the overt circularity from the use of MC in an NST proof; however, it leaves NS and the PPC suspended in midair on a postulate the main motivation of which is that it entails those results. Furthermore, it is questionable that merely dropping an explicit avowal of NS would wash away the *taint* of circularity involved in arguing from any sort of local field theory to the NST, since implicit in the whole mathematical structure of any local field theory is an acceptance of the Minkowski picture of spacetime as a backdrop against which quantum fields play out their histories.

What has really only been accomplished—as Eberhard and Ross (1989) are much more careful to argue than Eberhard (1978)—is a proof of the following conditional:[8]

$$(QM\& \; MC) \rightarrow NS \tag{5.6}$$

or, equivalently,

$$MC \rightarrow (QM \rightarrow NS). \tag{5.7}$$

In fact, we can say a little more than this. As Pauli guessed, and as we have shown in the previous chapter, given generally accepted rules of QM as background conditions, MC is a sufficient and *necessary* condition for NS:

$$QM \vdash MC \iff NS. \tag{5.8}$$

What we do *not* have, however, is the thing that would establish PPC with the generality it needs to serve as a philosophical principle, i.e.,

$$QM \vdash NS. \tag{5.9}$$

Physicists of the last generation were faced with the pressing need to get on with finding ways of making useful calculations, and they found it expedient to make what seemed to be a reasonable assumption that would guarantee that their theories would not conflict with relativistic causality. (Indeed, neither they had, nor we have, any clear idea what would be a more general rule with which one might *replace* MC.) So long as one is prepared to openly acknowledge that this is an expedient and somewhat provisional assumption there is no possible objection to it. But we risk falling into serious error if we try to take MC from QFT and use it somewhere else without keeping in mind its origins. The fact remains that in most versions of QFT commonly used today, MC is a *postulate* introduced to ensure the agreement of QFT with relativistic causality; *therefore any argument which claims on the basis of these versions of QFT that quantum theory does not conflict with relativity is manifestly circular.* The title of Eberhard & Ross (1989) is "Quantum Field Theory does not allow for faster-than-light communication"; of course

[8] I am grateful to Dr. Eberhard for clarifying this point for me (private communication).

not, it was *designed* not to. Therefore the confidence in the NST as an *unconditional* property of quantal systems seems quite misplaced.

Perhaps the problem is that no one has yet bothered to put the *whole* argument for peaceful coexistence together into one connected sequence of statements, all the way from Pauli's structuring of QFT to Shimony's conclusion that peaceful coexistence holds. Pauli's stricture that spacelike separate observables must always commute seems quite reasonable, given a prior acceptance of NS; Eberhard's confirmation that Pauli's guess is indeed necessary and sufficient to prevent signalling is, in itself, valid; Shimony's argument that we should accept the NST proofs as sufficient grounds for peaceful coexistence, given his stipulation that we have peaceful coexistence if and only if we have no signalling, and given that we can accept those proofs on authority, is reasonable in itself; however, conjoin all the premisses of all these arguments, and we have just shown that

$$NS \vdash NS, \tag{5.10}$$

which is valid, but not terribly informative. It seems to be unfortunately all too easy to forget the historical or logical antecedents of a clause of a theory and just use that clause *on authority*. (To be fair, even if we take this very rigorous view of the matter, Eberhard's proof and others like it are quite useful at least to the extent that they demonstrate the internal consistency of the theory and help to illuminate its mathematical structure.)

5.3 Limitations of MC

What, then, if we adopt the suggestion given above and use a formal or semi-formal version of QFT which includes MC merely as an unjustified postulate? Then the very interesting problem becomes to establish the scope and limitations of MC—which then become the scope and limitations of peaceful coexistence (at least in the narrow sense we have so far indicated).

In view of the philosophical stake we have invested in MC as a basis for proofs of the NST, it is important to realize that experts in field theory have often spoken of the limitations of this hypothesis. Many of the classic texts in QFT—such as Streater and Wightman (1964) and Bjorken and Drell (1965)—show a great sensitivity to the possible breakdown of MC. Bogolubov *et al.* (1975: 508) have this to say, for instance:

> We have already pointed out that the axiom of local commutativity is one of the most restrictive principles of quantum field theory. One might question whether this axiom is well-founded experimentally. We do not have adequate grounds for asserting that measurement of the components of an Hermitian field at some point does not influence the value of the components of this field at another point at a spacelike distance of the order of 10^{-16}cm (or less). . . . attempts to introduct 'nonlocality in the small' are often accompanied by the assumption of some anisotropy of the vacuum or a violation of the customary strong formulation of relativistic invariance.

Some further remarks of Schweber's also help to put the matter in perspective:

> It is [the] inability to give convincing and mathematically rigorous proofs for many of the assertions made on the basis of the quantum theory of fields. . . which has been the principal motivation for the important investigations of the general structure of local field theories which have been carried out in recent years.. . . Their principal aim is to discover whether any local relativistic field theories exist. [!!] Their approach has mainly consisted in studying the consequences for observable quantities of the locality assumption, i.e., that the commutator (for integer spin fields) or the anticommutator (for odd half-integer spin fields) of two local field operators vanishes for spacelike separation.. . . [1961: 721]
>
> Whether physical particles and their interactions can be accounted for by a description in terms of local operators remains at present an open question. [1961: 723]

We can take two things from Schweber's comment. First, MC is sufficiently problematic that if peaceful coexistence must depend upon it alone then peaceful coexistence rests upon rather uncertain ground. Second, so-called axiomatic local quantum field theory can be understood as a *project* to see how much progress can be made under assumptions which guarantee as little conflict with relativistic causality as possible; it cannot be taken as a theory which somehow justifies relativistic causality from less questionable prior principles.

One might point out that Schweber's book is now thirty years out of date, and that many of the limitations of QFT that Schweber speaks of may have since been overcome. This may be true, although a survey of more recent texts does not suggest any fundamental progress in our understanding of causality.[9] However, it is interesting that the recent (1989) review article by Eberhard and Ross uses Schweber's book at its primary reference on field theory, and openly takes local field theory as paradigmatic of field theory. Therefore, what Schweber and Pauli have to say is not of merely historical interest, but is certainly relevant to the current debate.

It is at least thinkable that MC might break down under those exotic circumstances which would require the long-sought theory of quantum gravity for their analysis—perhaps in extreme gravitational fields, in the interior of black holes or the early universe. In fact, my thinking about the possible limitations of the NST was crystallized by certain remarks along these lines made by Roger Penrose, in a lecture at the University of Toronto, on September 27, 1990. Penrose pointed out that when and if a true quantum theory of spacetime becomes available the NST might break down *via* a failure of microcausality, since there might be circumstances under which spacetime separation itself is not definable. Note that this could be a problem which would occur not merely at very short distances, as Bogolubov *et al.* speculate.

[9] I readily concede that I am not expert on recent field or particle theory, and may very well be unaware of current developments. But unless I am quite mistaken recent progress has largely been within the general framework of local field theory. See Lee (1981), for instance.

I do not deny that a failure of MC would have enormous theoretical consequences. Perhaps the most serious of these might be the failure of the spin-statistics theorem, which depends essentially upon MC.[10] Since recent experiments have verified the exclusion principle to a high degree of accuracy[11] it seems quite unlikely that the spin-statistics theorem—and therefore MC and probably also the NST—fails for any kind of physics that we might be normally familiar with. However, it was not the purpose of this inquiry to suggest otherwise. My purpose has been to suggest that at least some of the attempts to prove the NST *in all generality* stand on a much less secure logical footing than is commonly appreciated. We are concerned with a question of principle here. The fact that we cannot bring MC into physics except by assumption should alert us to the possibility that it has limitations, in spite of the fact that we presently do not know how to do without it. Or perhaps we could put this a little more optimistically: the fact that MC and the NST clearly do hold for a large part, at least, of the physics we are familiar with suggests that when we do learn how to justify them some quite exciting new physics might thereby be exposed.[12]

5.4 MC and Relativistic Covariance

We have already seen, in the last chapter, that a breakdown of MC would, at least in principle, permit superluminal signalling. It is important to see, however, that other kinds of arguments (subtly but importantly flawed, I will eventually argue) can also be given to show that MC is required to make quantum field theory relativistically consistent. I have never seen the following arguments spelled out in full, perhaps because they are considered too obvious. (It is not uncommon for a scientific or philosophical debate to turn on the points that everyone thought were too obvious to bother mentioning.) In fact, I cannot find in the literature any explicit justification of MC in full generality that does not simply amount to a statement that we have to have MC *because otherwise signalling would be possible.* Nevertheless, something like the following arguments seems to be *implicit* in the general view of MC. As we shall eventually see, trying to answer these arguments will eventually lead us to grounds for a truly adequate description of the justification *and possible limitations* of MC.

First, let us outline an argument to show that if MC failed, the state of a system would be frame-dependent. Suppose there are certain operators A and B which represent measurements made on spacelike separate parts of a spatially extended system. Let us suppose also that these observables do not commute; that is, we shall assume that there is some other operator C such that

$$[A, B] = C, \qquad C \neq 0. \tag{5.11}$$

(C would not in general be Hermitian and therefore would not in general be an observable itself.) What the commutation relation above says is that the order in which the

[10] See Bjorken and Drell (1965), p. 170–172, or Schweber (1961), p. 222-223.

[11] Kekez *et al.*, (1990).

[12] I am indebted to J. R. Brown for pointing out to me that this somewhat more encouraging viewpoint is an option.

measurements A and B are made makes a difference to the global state of the system (and therefore presumably to probabilities and expectation values). If A is performed first and then B, the system in general goes into a different state than if B is performed before A. One can immediately see that if this can indeed hold for measurements performed at a spacelike separation there is a severe problem finding a relativistically consistent description of this process. For there exists a reference frame in which A occurs before B and a reference frame in which B occurs before A; therefore it seems as if the state of the system depends upon the choice of reference frame. Even if we don't agree that the abstract notion of the system's state should be covariant—and there are considerable difficulties with the concept of a covariant description of the state vector (see Aharonov and Albert 1980, 1981)—it certainly would seem that the probabilities and expectation values that we can calculate for the system ought not to depend upon the reference frame; and yet these are functions of the system's state. The only way we could get out of this would be to impose a global or intrinsic time ordering of some sort, so that regardless of appearances in different frames there would be some invariant or intrinsic sense in which A occurs before B.[13] Bell has suggested that it might not be utterly destructive to the Principle of Relativity, properly understood, if we were forced by quantum nonlocality to reintroduce some sort of absolute reference frame or quantum ether to physics (Davies & Brown 1986), but there is no question that this would be a gross violation of *textbook* relativity.

We can also make a simple but very powerful argument for MC from the relativity of simultaneity. QM tells us that two measurements can be made simultaneously if and only if they are compatible (i.e., commutable); however, any two spacelike related points x and x' are simultaneous in some frame. Therefore, any measurements made at x and x' must commute. This is such a good argument in favour of MC that it might be thought to settle the question; however, I shall suggest in a later chapter that does not quite work as we might think because of a crucial ambiguity in the notion of simultaneity.

Note that the arguments in this section do not *establish* peaceful coexistence but *presume* it.

5.5 Summary

If we don't justify MC by appealing to the localizability of observables, then how do we justify it? There are three possibilities:

1. Introduce it as a primitive axiom; then the no-signalling argument is no longer circular, but does not have the generality to support the PPC. If we can do no better than treat MC as a primitive then we might as well concede that the PPC

[13]Actually there is another thinkable escape route. If the commutation relations for spacelike separate operators were such that the hypothetical operator C in the above equation turned out merely to be multiplication by some phase factor, the choice of reference frame would of course make no difference to the observable expectation values and probabilities that we could calculate. Unfortunately, we know nothing about what the detailed form of some nonzero commutator like Equation (5.1) might be.

is unfounded, or at least that we do not understand why we are apparently compelled to believe it.

2. Appeal to Pauli's reasoning (which is that there is no signalling, and since a failure of MC would allow signalling, MC must hold); then the argument that uses the Eberhard proof (or any other which justifies commutativity of operators on the basis of MC) to justify the PPC is manifestly circular. Note that Pauli makes no explicit appeal to the locality of measurements; he just points out (and Eberhard confirms) that a failure of MC would violate relativity.

3. Appeal to the relativity of simultaneity; argue that since all points outside the light cone are equivalent in this respect (i.e., can be simultaneous in some frame), and since measurements that can be made simultaneously are compatible, MC must hold. This is in many ways the strongest argument in favour of MC, although it does not seem to have been often explicitly stated in the literature. I will return to the central question of simultaneity later on.

Chapter 6

Where Do We Go From Here?

> For me then this is the real problem with quantum theory: the apparently essential conflict between any sharp formulation and fundamental relativity.... It may be that a real synthesis of quantum and relativity theories requires not just technical developments but radical conceptual renewal.
>
> — J. S. BELL (1986: 8, reprinted in 1987b)

> It appears, from the point of view of quantum mechanics, that the theory of relativity lies at the same level as classical mechanics in that it only represents an approximation relative to the macroscopic domain. One will not just admit relativity formulas (for instance, Lorentz formulas) to be valid without change in the intra-atomic domain. They will have to suffer modifications which will be probably analogous to the ones suffered by ordinary mechanics to transform into quantum mechanics. We have to "quantize" the Lorentz transformation.
>
> — E. SCHRÖDINGER (1984, quoted in Bacry 1988: 79)

We have gone a long way in this thesis to arrive at a very simple conclusion: all the No-Signalling proofs that have been published so far either implicitly or *explicitly* assume that measurements have only local effects; therefore, they certainly do not *prove* that measurements have only local effects, which is precisely the thing that is of interest in a defence of the PPC.

The difficulty inherent in the presently accepted view shows up quite clearly in this statement by Stapp (1988: 376):

> It must be strongly emphasized that EPR locality ["a mere change in the choice of the experiment performed in one region cannot disturb events in a spatially-separated region": 371] is not an actual consequence of the theory

> of relativity. The basic demand of the theory of relativity is covariance, which pertains, strictly speaking, to deterministic systems that are well defined over all of spacetime. The failure of EPR locality does not jeopardize covariance. Nor does it jeopardize relativistic quantum field theory, *which is constructed to ensure* that its predictions do not depend either on the frame of reference or upon the order in which one imagines performing measurements on spacelike separated regions. Moreover, in spite of the failure of EPR locality, no signal can be sent faster than light by any system adequately described by quantum theory.
>
> This last-mentioned property arises from the fact that the quantum-theoretical *probabilities* in one region are independent of the choice of experiment made in a spatially separated region.

I find it very hard to understand why Stapp and others do not see that the proof of a result on the basis of a theory which was "constructed to ensure" that result is no proof at all. As I have noted, all the proof shows is that the construction was successful: it accomplishes what it was intended to accomplish. The authors of the various proofs we have examined are not careless thinkers; Eberhard, indeed, is extraordinarily thorough. (His (1978) is an impressively painstaking piece of analysis.) They can only have fallen into the trap of circularity either because of a failure to thoroughly examine the background assumptions of the theory they were using (as in the case of the QFT-based proofs) or because certain assumptions about the localizability of observables seemed so reasonable and obvious that they did not seem to merit critical scrutiny.

Indeed, the debate, if there is to be a debate, will be about what we are entitled to assume in an inquiry of this nature. Maybe one of the most useful lessons we can take from this story is just a reminder—yet another reminder, one might say—that it is enormously difficult in a discussion of deep principle to avoid assuming the very thing one needs to prove, especially when there is a great deal riding on the outcome. (The many attempts to found the Axiom of Infinity throughout the history of mathematical logic are perhaps another testament to this fact).

Where do we go from here? The only answer, of course, is to carefully rethink the whole question of the relation of relativity and QM, paying particular attention to the analysis of those concepts which up to now have generally been considered entirely unproblematic. I will briefly sketch a few considerations of methodology and principle which might be useful in this endeavour.

6.1 Methodological Considerations, or, "Get a Good Book. . . "

In his (1990), the late John Bell underscored, with gentle wit and a wisdom that could only have been founded in a lifetime of thought and study, the fact that the quantum measurement problem is still unsolved. Nevertheless, he entirely agreed that ordinary quantum mechanics is just fine "FOR ALL PRACTICAL PURPOSES", and he said that this defensive phrase crops up so often in such discussions that it is handy to have an abbreviation for it: FAPP. Let us indeed stipulate, with a qualification to be mentioned in a moment, that ordinary quantum mechanics *and* ordinary relativity theory are just fine FAPP.

The problem with this soothing phrase is that it presumes we know what the word “practical” means. (In fact, Bell ironically characterized the kind of reasoning that relies on FAPP to exit from a difficult problem as a “fuzzy logic”; he was well aware that in the last analysis this phrase is just an evasion.) “Practical purposes” are purposes that we are *presently* concerned with; they are defined in terms of problems that we presently face or opportunities and possibilities that we are presently aware of. However, there can be no question that our conception of what is practical will continue to evolve as it has always evolved, perhaps at times more rapidly than we might find to be comfortable. The “practical purposes” that can be accomplished with a miniaturized solid-state laser, for instance, (such as writing data to ultra-compact storage devices) could only have been science fiction fifty years ago.

It is true that in an investigation like this we are concerned with questions of principle. But questions of principle have a way of quite rapidly turning into questions of practicality. I am sure that a historical study would show that there are fewer more powerful means of changing the notion of the practical than attempts to resolve theoretical questions of principle, so long as they are genuine attempts to wrestle with basic conceptual problems and not just scholastic hair-splitting. (Recall, for instance, that Einstein created special relativity primarily in order to resolve the theoretical mismatch between Maxwellian electrodynamics and Newtonian mechanics.) Therefore I do not think that there is any justification for supposing that probing the limits and boundaries of our best theories will never be anything more than an idle philosophic amusement. In fact, given our present state of global ecological, political and intellectual crisis, one might well attach some urgency to the resolution of a very fundamental hiatus in our understanding of natural process.

The point of these remarks is that in the (not very) long run nothing could be more *impractical* than to dismiss a difficulty that seems to have no *immediate* practical bearing. The obscure theoretical puzzles, the paradoxes, are precisely the loose blocks in the wall that bars us from the practical solutions we need to find (even if we do not yet know that *those* practical solutions are the ones we need).

Bell also cautioned us against the advice that all we have to do is “get a good book”, as if just doing a little homework is all we need to see our way through the problems of measurement and nonlocality. Not that we should not do our homework; far from it! But a large part of the mistake that has been made in the discussion of the PPC (I will not say the debate, because there has not been enough debate) is to assume that there is simply one, canonically correct and universally accepted formulation of quantum mechanics. This is certainly false; there are many versions, almost as many versions as there are authors and many versions which differ significantly either in their philosophical underpinnings or in their mathematical formulation or both. (See Gudder (1977) for a detailed outline of four quite different attempts to formalize QM, for instance.) And versions of QM (relativistic or otherwise) often tend to differ precisely on those points which are crucial in the discussion of the PPC, such as the nature of the measurement process, the nature and interpretation of state change or reduction, or the formulation and justification of commutation rules. The day has not yet come when the treatment of the signalling question is a textbook exercise for second-year students.

Another methodological consideration is simply to be on guard against dogmatism and complacency. It is all to easy for someone who is "expert" in some complex field to fall into this trap. At the head of this chapter I quoted certain eminently perceptive and wise remarks by J. S. Bell. In the same piece, however, Bell himself says some things which strike me as excessively self-congratulatory—but which would not be worth mentioning were it not that they are so typical. "As for technical mistakes," he says (Bell 1986: 7), "our theorists do not make them." Indeed! I suppose it depends upon what you mean by a "technical" mistake. Perhaps there are very few errors of calculation or simple fact, typos, and so forth that escape today's flinty-eyed referees and technical proof readers. But how do we know that we are not all making some error that future generations will see as simply *foolish*? After all, it was Bell himself (1966) who caught out the normally unimpeachable John von Neumann in an error that Bell himself later characterized as "foolish" (1988). One can just see a physics text of 2091, written in that familiar breezy know-it-all style: "Of course, we all know now that those guys back in 1991 had no need whatever for the concept of. . . " And did Bell really believe that "our theorists" have no "obsessive committment" to hypotheses, no "religious intensity"? Is it really true that our age has mounted some unique plateau of calm objectivity? I find this very unlikely. . .

6.2 The Unfinished Revolution

Throughout this study, I have tended to characterize the No-Signalling Theorem as an attempt to protect relativity theory from the inroads of nonlocality. But in fact many authors of these proofs (Eberhard 1978, Ghirardi *et al.* 1980, Eberhard & Ross 1989) seem to see themselves as protecting *quantum mechanics* against the charge that it might permit violations of relativity. I would like to suggest that this concern for the integrity of quantum mechanics is somewhat misplaced. While there clearly are problems with QM—particularly, as noted above, in what might be called its auxiliary parts such as measurement theory—it is much more fundamental than any classical theory, including (I suggest) relativity theory. There are examples too numerous to mention of cases in which some quantum rule may be found to hold when the corresponding classical rule breaks down. We have ample reason to believe that most if not all classical laws are limiting cases of quantum laws. In all likelihood, therefore, the classical worldview—presumably *including* relativity, which is simply the farthest extension of the classical picture—is merely some sort of limiting case of the quantum worldview. Of course, this must not be regarded as *proven*—only when a thorough reduction of all classical concepts to quantum concepts has actually been carried out could we say that it has been proven—but the overall trend of physics in this century certainly makes it seem very probable. It seems very likely that relativity, like all other classical theories, will eventually be found to be an approximation to some deeper (and stranger) quantum theory.

It is worth quoting Misner *et al.* (1973: 1182–1183) on the ultimate breakdown of spacetime:

> The uncertainty principle. . . deprives one of any way whatsoever to predict. . . "the deterministic classical history of space evolving in time". *No prediction of spacetime, therefore no meaning for spacetime,* is the verdict of the quantum principle. That object which is central to all of classical general relativity, the four-dimensional spacetime geometry, simply does not exist, except in a classical approximation.
>
> These considerations reveal that the concepts of spacetime and time are not primary but secondary ideas in the structure of physical theory. These concepts are valid in the classical approximation. However, they have neither meaning nor application under circumstances where quantum geometrodynamic effects become important. Then one has to forego that view of nature in which every event, past, present, or future, occupies its preordained position in a grant catalog called "spacetime", with the Einstein interval from each event to its neighbour eternally established.

Of course, Misner *et al.* may be wrong about this, but there are such very good reasons for believing them to be more or less right that I find it rather surprising that so many authors have espoused the notion of peaceful coexistence with such confidence. The whole trend of physics in this century seems to rather obviously show that the ultimate breakdown of "peaceful coexistence" is *exactly* what we should expect.

In spite of this, however, in ordinary quantum theory and especially in quantum field theory we assume the relativistic picture of spacetime as given from the outset. In QFT fields and particles play out their histories against a backdrop of Minkowski space. In fact, quantum field theory represents a step *backward* philosophically from ordinary nonrelativistic quantum theory, since it treats *both* position and time coordinates as *c*-numbers (i.e., continuous commutable parameters). We have to assume that this is a temporary state of affairs, to be rectified when a genuine quantum theory of spacetime becomes available. Present-day local QFT should probably be viewed as a semi-classical approximation, perhaps analogous to (although enormously more sophisticated than) the Old Quantum Theory. The quantum revolution that Max Planck somewhat unwittingly triggered in the last weeks of the 19th Century is still unfinished; it will not be finished (if it ever can be) until we have a genuinely quantal theory of space and time.

6.3 Signalling in Non-Standard Quantum Theories

It is worth mentioning that the possibility of superluminal signalling has been acknowledged to exist in certain non-standard versions of QM. Eberhard (1989b) has published a version of QM in which distant correlations are established not instantaneously but at some definite superluminal velocity. Eberhard showed that if this velocity is made high enough, the predictions of his theory can be made arbitrarily close to those of standard quantum theory; however, he also found that signalling would be possible in his theory. Also, Polchinski (1991) has pointed out that superluminal signalling is possible in a non-standard version of QM put forward by Weinberg; in Weinberg's theory, Schrödinger's equation is modified into a nonlinear form in an attempt to replace the discontinuous process of state reduction in standard theory by a continuous process. Whether these results should be considered *reductii ad absurdum* of these tentative theories remains to be seen.[1]

6.4 Is Space Fundamental?

Newton essentially believed that the notion of any kind of existence *not* in space is simply incoherent; existence is existence in space. Newton's reasons for believing this are complicated and subtle, and I will not attempt to analyse them here, save to note that he was obviously right at least to the limited extent that it is certainly very hard to *picture* something in the imagination without placing it in a background. Leibniz, on the other hand (and Kant after him, in a different way) believed that space and time are only orderings of appearances or phenomena, and that being at the deepest level is not spatially organized.

The quantum theoretical view is much more in accord with Leibniz than Newton. QM says that there is nothing fundamental about space (and it would say this about time as well, if we knew how to represent time as a quantum operator). Chester puts this very clearly:

> Quantum mechanics teaches us to view the universe as measurement spaces, not physical position space. Position space is not sacred among measurement spaces; it is only a profane one among many of them.
>
> Only if it is in a state of position does something have a position.[2]

We could roughly, but not entirely inaccurately, characterize the development of quantum theory in the 20th Century as deciding the Leibniz/Newton debate in favour of Leibniz, although in a way that he would have likely found shocking.

If space is not fundamental, what about spacetime? In relativity theory, space and time are united into the invariant *interval*:

$$s^2 = c^2t^2 - x^2 - y^2 - z^2. \tag{6.1}$$

If space is not fundamental, then spacetime cannot be either. However, we presently have no way of representing interval as an observable or operator, because we have no way of representing *time* as an operator. There is much disagreement about whether or not this problem is soluble, with the orthodox view being that it is not. It is certainly beyond the scope of this thesis to examine this very difficult problem. (See Recami

[1] I thank Dr. Eberhard for bringing this work to my attention.

[2] Chester 1987, p. 307.

1977, Prigogine 1980, and Schommers 1989 for discussion and some suggestions for a solution.) My own hunch is that the problem *must* be soluble (perhaps by restricting possible times to a discrete spectrum); the fact that we cannot make progress on such a fundamental issue suggests that our notions of time need careful rethinking. (In the next chapter I will sketch some ways in which this might be done.)

A genuine proof of the NST in the context of a quantum theory of spacetime would involve the following: it must be possible to write *both* space and time coordinates as operators (not just the position operator); this would allow us to define an operator for spacetime interval. Systems would then be *in spacetime* only when they are written in this representation, just as we say that systems have a position only when they are written in position representation (Chester). The question of signalling is the question of signalling with respect to spacelike separation; therefore we would have to show whether or not there can be signalling between branches of a composite system in which the branches are separated with respect to interval. Hence a truly adequate treatment of the question of signalling in nonlocal quantum systems requires a full-fledged quantum theory of space and time—something that, needless to say, we do not yet have. In the next chapter I will make some tentative suggestions that could, perhaps, help to clear the field for the construction of such a theory.

Chapter 7

A Spacetime Description of Nonlocality

> Today we say that the law of relativity is supposed to be true at all energies, but someday somebody may come along and say how stupid we were. We do not know where we are "stupid" until we "stick our neck out," and so the whole idea is to put our neck out. And the only way to find out that we are wrong is to find out *what* our predictions are. It is absolutely necessary to make constructs.
>
> — RICHARD P. FEYNMAN, *The Feynman Lectures in Physics* (1965) p. 2-9.

> It is one thing, to shew a Man that he is in an Error, and another, to put him in possession of Truth.
>
> — JOHN LOCKE, *An Essay Concerning Human Understanding* (1690)[1]

This dissertation has, up to this point, been largely critical. I have really just being pointing out that not enough attention has been paid to certain things that are or should be quite well known, such as the fact that QFT was quite deliberately constructed by its founders in such a way as to guarantee peaceful coexistence. I have really said very little in a positive direction that is original. However, I would not want it thought that I was content merely to play the rôle of the obstreperous little boy in "The Emperor's New Clothing"; in this final chapter, therefore, I would like to offer some constructive suggestions as to how we might get out of the conceptual logjam that is presented to us by quantum nonlocality. The material in this chapter is necessarily conjectural and therefore much more tentative than the main body of the argument offered so far. However, even if the conceptions I sketch here are not correct I believe that if they are taken seriously they may point in a useful direction of inquiry.

[1] Quoted in Knuth (1986: 303).

The question I will examine here is essentially this: how can we describe the totally nonlocal process of state reduction in a relativistically covariant way? This is arguably one of the central questions of modern natural philosophy, and has an important and interesting bearing on our related problems of signalling and peaceful coexistence.

The main line of argument I have been taking in this thesis is that the no-sigalling proofs so far published do not really succeed in establishing the NST because they are either circular or else dependent upon such excessively strong assumptions about the localizability of measurement interactions as to rule out *ab initio* the very cases they should be equipped to consider. However, I also pointed out (Chapter 5) that there actually is a very convincing, although seldom clearly stated, argument for microcausality from considerations about the relativity of simultaneity. (All observations that *can* be simultaneous must commute; all spacelike separate points can be simultaneous in some frame; therefore all possible observables acting at points outside each other's lightcones must commute.) A defender of peaceful coexistence might well say this: "Very well, I concede that some of us have cut a few logical corners in proving the NST; however, the argument that you yourself have advanced in favour of MC certainly holds for any spacetime coherent enough to possess a light cone structure; there certainly are such spacetimes since the reality of distant EPR correlations does not contradict the fact that the speed of light is an invariant; therefore (by the Eberhard proof, the GRW proof, or any other proof depending upon commutativity) the NST holds for normal spacetimes and peaceful coexistence, at least For All Practical Purposes, is saved. Therefore this whole discussion is not very interesting, anyway. . . "

I could just respond to this by saying that we are interested in questions of principle here, so the question as to what merely holds for "normal" spacetimes is not decisive. But a much stronger and more interesting sort of reply can be made to this objection. In this final chapter I will try to show that if we (i) describe the state reduction process in spacetime in a certain very natural, but unconventional way, and (ii) carefully rethink the notion of simultaneity, my seemingly transparent argument for MC does not go through quite as outlined above, and that in fact there will appear certain natural *limitations* upon MC even in ordinary, causally well-behaved spacetime.

7.1 The Problem: Is State Reduction Instantaneous?

In order to discuss what goes on, or what seems to go on, in experiments of the EPR type, let us consider a simple, if not simplistic, spacetime diagram of a typical EPR set-up. (See Figure 7.1.) For definiteness, we can suppose that the two-particle system is in a singlet state.

S is the particle source, and for convenience we will, as usual, label the particles as "left" and "right". Let A be the world-point at which the left particle encounters a detector. Acording to the usual way of describing such a system, the left particle is not in a definite spin state in the portion of its life from S to A. We know that there is only a probability of $1/2$ that this detector will find the left particle to possess spin in (say) the z direction. However, afterwards (in terms of proper time along the left particle's worldline) we say that the particle possesses a definite S_z state. This statement has the precise (and somewhat narrow) meaning that it is possible to carry out a *quantum non-demolition measurement* or *monitoring experiment*[2] which will always find the left particle to be in that same spin state. A *non-demolition measurement* (or *monitoring experiment*) is defined as follows (Braginsky, Vorontsov, & Thorne 1980, in Wheeler & Zurek 1983: 753; see also Aharonov & Albert 1980, 1981):

> We define a QND [quantum nondemolition measurement] of [an observable] $\hat{A}$ as a sequence of precise measurements of $\hat{A}$ such that the result of each measurement is completely predictable from the result of the first measurement...

The appropriate QND in this case would be to just keep on measuring S_z. If the left particle is allowed to pass through a series of detectors they will record the same value of S_z for the particle as was recorded by the first detector at event A. (I am assuming that the detectors are designed in such a way as to not absorb the particle in the act of detecting it. As far as I know, this poses no difficulty *in principle*. Such a detector might just be a device which absorbs all particles which are *not* deflected in a certain precise direction by a certain arrangement of magnets.)

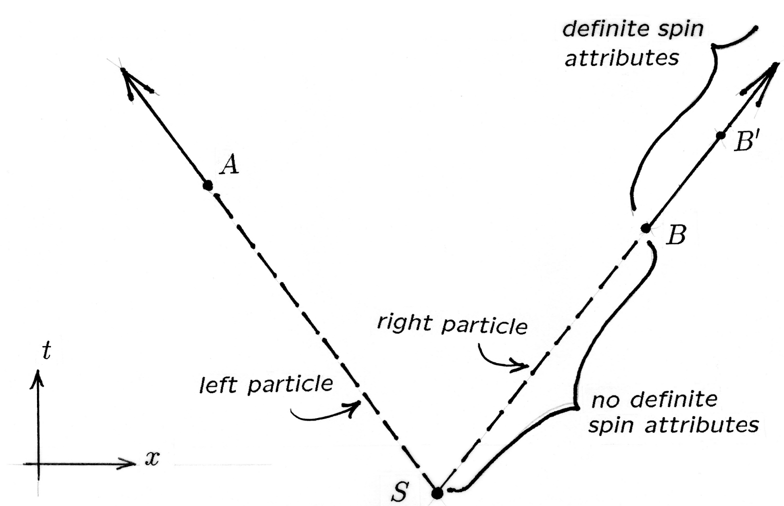

Figure 7.1: A Typical EPR Scenario

[2]There is some variation in terminology here. Aharonov & Albert 1980, 1981 define a non-demolition experiment as a single measurement which is simply a repeat of a measurement just performed on a system; since the first measurement prepared the system in a certain state, the second measurement simply verifies that the system is in that state and can be repeated indefinitely without further changing the state of the system. They call a sequence of such experiments a monitoring experiment, which is the same thing that Braginsky *et al.* call a QND.

Let us call the event A in the worldline of the left particle an *amplification event*, following Bohr, who suggested that when measurement drives a quantum system into a classically definite state that an "irreversible act of amplification" occurs.[3] This is the event at which the particle becomes "classical", or at least at which its attributes become definite in the precise sense indicated above.

What about the right particle? We know that there is certainly some event B' after which (in terms of proper time along the right particle's world line) a monitoring experiment will find the right particle to possess a definite S_z state (in fact, the opposite of the one found for the left particle). However, we also know that we get into a contradiction with experiment (with experiments performed on ensembles of particles prepared similarly to these two particles) if we assume that the right particle was in a definite spin state all the way back to S; I suggest that there has to be some *particular* event B on the right worldline at which the right particle acquires a definite value of S_z. This would also be an amplification event; however, it is *not* necessarily the world point at which the right particle first encounters a detector, if it ever does. We know that only one of the two particles need encounter a localized detector in order for the global state of the system to change.

It is, of course, unclear just what is getting amplified in this case, since the right particle does not necessarily encounter a Stern-Gerlach detector, spark chamber, or other such device. However, since there has to be a definite spacetime point at which the particle *becomes* classical there has to be, if Bohr is right, some sort of amplification process at that point whether or not there is a detector there in the usual sense. My conjecture is that space or the vacuum itself—which as Lee (1981: 826) suggests should be thought of as a "physical medium"—acts as if it were a sort of grand spark chamber or 3-dimensional photoemulsion; a classically definite particle trajectory is a sequence of particle registrations on the fabric of space itself, each registration being an amplified excitation of the vacuum. It goes without saying that this notion needs some fleshing out, but something of this kind has to be the case.

One might be tempted to argue that the right particle becomes definite only when it encounters a detector. To see that this is false, suppose that the right particle *never* happens to encounter a detector, and consider what the experiment looks like at some very late time in the frame of reference of an observer at rest with respect to the source. (There's nothing special about this frame; I just pick it as a convenient overall reference frame.) The left particle was found in a definite spin state; therefore there is certainly some late enough time after which we would get a violation of the conservation of

---

[3] Wheeler, "Law Without Law", in Wheeler & Zurek 1983:

> For Bohr the central point. . . [is] an experimental device. . . capable of an "irreversible act of amplification".

See also Bohr (1963: 92, 1958: 73, 88). From (1958: 73) we have

> . . . every atomic phenomenon is closed in the sense that its observation is based on registrations obtained by means of suitable amplification devices with irreversible functioning. . .

spin if the right particle could not be said to have the opposite spin value. Therefore, consistency with conservation laws demands that there be some point B at which the particle does, indeed, become individuated or acquire definite properties, whether or not the particle is ever observed. We might say that B is the point at which we acquire the right to speak of the particle counterfactually; we can say that if a measurement of S_z *were to be made* on the particle after B (where, as usual, "after" means in terms of proper time along that particle's worldline) a certain value would be found with probability one.

Note that there are obvious problems in trying to decide whether or not we should say that A "causes" B, since their temporal order could be different in different frames of reference. This is an important problem, but we will sidestep it for now. Our main concern will be with *identifying* the point B, and, especially, determining whether its location is frame-dependent.

We could perhaps speak of the matter this way also: *given* that the left particle interacted[4] with the left detector at world-point A, there is a point B on the right worldline after which measurement of the *right* particle could make no difference to measurements on the left particle; or more precisely, measurements on ensembles of similarly prepared particles will show that there are two definite points A, B on the left and right worldlines respectively after which the measurement results will no longer exhibit the nonclassical correlations responsible for the violation of the Bell Inequality in such experiments.

Normally, in nonrelativistic QM, we identify the points at which the system undergoes some sort of irreversible amplification due to measurement as the points at which the state vector reduces to, or is projected onto, a definite eigensubspace of the system. If the notion of state reduction means anything at all (and of course there are those who argue that it does not, such as Ballentine 1990a,b), then for the type of case we consider here, the points A and B should also be thought of as the points at which the state of the total system reduces. State reduction is, of course, unobservable in itself, but it seems that we must say that there are distinguished points such as our A and B at which the left and right particles become uncorrelated, and according to our usual theoretical description these are the points at which the system reduces.

Now, it seems intuitively reasonable to think that it must be an *invariant* fact that the left and right particles become uncorrelated at two particular points on the left and right worldlines, given that a measurement took place on one of the pair of particles at a certain point on its worldline. However, this does not accord at all with the usual way of speaking of state reduction, which is that it is *simultaneous in the frame of reference of the observer*.[5] This would mean that in the scenario that we have just sketched here, the amplification event B would *necessarily* be simultaneous with the event A in the frame

[4] Interacts? One does not quite know what tense to use here!

[5] For instance, Penrose (1989: 480) says, "... an 'observation' ... at one end of a room can effect the *simultaneous* reduction of the state-vector at the other end". See also Figure 6.32 in Penrose (1989), which shows the "state 'jumps' " as supposedly happening in the simultaneous space of each observer, and Aharonov & Albert, (1980, 1981).

of reference of the observer that interacted with the left particle at A. A few simple considerations should show that this is not an especially coherent notion.

First, let us note that it is an invariant fact that the worldline of some detector was coincident with the worldline of the left particle at A. The location of A in spacetime is something that all observers will agree on. What about B? Suppose that B was, indeed, determined by the intersection of the left observer's plane of simultaneity with the right particle's worldline. The problem is that we know nothing about the state of motion of the left observer. All that matters in our scenario is that his detector somehow intersects the worldline of the left particle; however, his plane of simultaneity could lie anywhere in the angle α in Figure 7.2. This means that given only the initial state of the system,

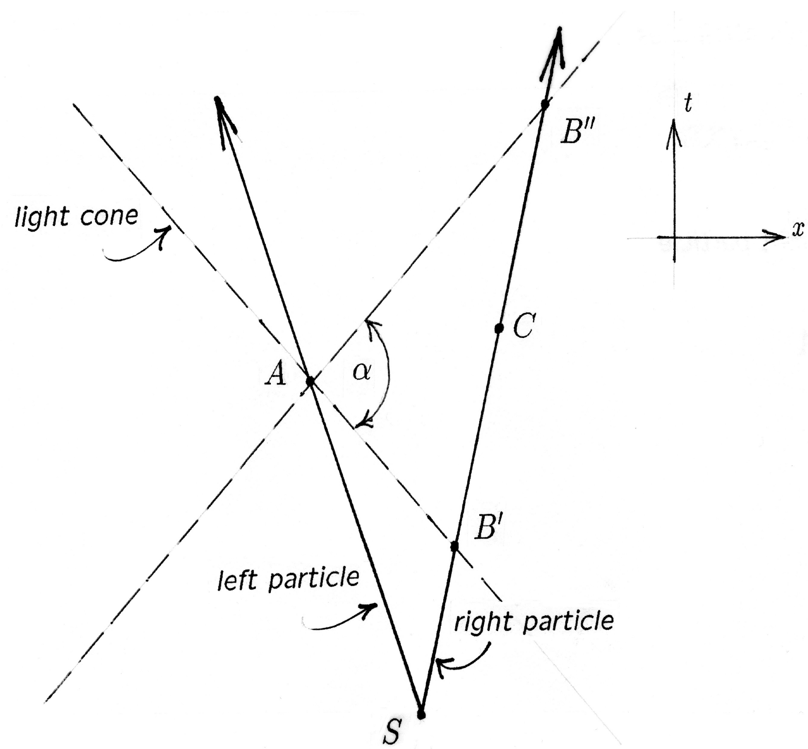

Figure 7.2: Where is B?

and the spacetime location of A, the amplification point B for the right particle could be anywhere along $B'B''$. In fact, if the left and right particles were travelling at almost the speed of light, and the left detector was moving in the negative x direction at almost the speed of light, B could be made very nearly coincident with S. In other words, the right particle would be in a definite spin state for nearly its whole history, merely because of an accident in the way its partner was observed.

Suppose now that the right particle were to be measured at an intermediate point

C by an observer moving in the positive x direction at nearly the speed of light. Let us say that this observer was aware of the circumstances under which the left particle was being measured (this could have all been arranged beforehand). The right observer's hyperplane of simultaneity could be made to intersect the left worldline very close to S; therefore the right observer (at least if she believes that state reduction is simultaneous in the frame of the observer) will conclude that the particle pair was not really in the singlet state at all, except for a vanishingly small interval at the beginning of its flight. Meanwhile, a third observer—perhaps one who was at rest with respect to the source—might say that the system was in a singlet state all the way up to the time in the third system at which the first measurement was made.

Would correlations between measurement results taken at C and A confirm, or disconfirm, that the particle was in a singlet state? (To find such correlations we would, of course, have to run the experiment over and over again.) If not, the state of the total system would have been determined by the state of motion of the detectors; thinkable, perhaps, but something for which there is certainly no provision in current theory. Bearing in mind that an experiment of this nature has never been performed, I think it is most reasonable to conclude that measurement results at the two world-points A and C would not be found to depend at all upon the relative velocities of the detectors which make these measurements. However, they certainly would, if the abstract process of state reduction—and the corresponding very real change in observable correlations—were a frame-dependent process.

One can construct other paradoxes of a similar nature. The following variant of the above (in fact the prototype of the above argument) is well-known (see Ghirardi, Grassi & Pearle 1990: 1300–1301). Consider some entangled two-particle state (Figure 7.3).

Assume that the particles are to be observed at points A and B, and assume, as usual, that we believe that state reduction is instantaneous in the plane of simultaneity of the observer. Suppose this system is viewed by two observers, O and O', moving relatively to each other in such a way that their planes of simultaneity are as shown. Then for O, A is earlier than B, while for O', A is later than B. O says that the system was already reduced at B, while O' says that the system was already reduced at A, even though it would seem that the state of the system ought not to depend merely upon the velocity of the frame of reference from which it is observed.

If we insist that state reduction is simultaneous in the frame of every observer that interacts or could interact with the system, we obviously cannot find a relativistically consistent description of the reduction process. It seems like the only way we could get this would be to assume that state reduction takes place in some definite, invariant hypersurface. In effect, we need a covariant notion of simultaneity—which, to those of us schooled in the relativistic way of thinking, seems like a contradiction in terms.

Would O and O' agree on their predictions for the system? This turns out to be a very subtle question, an answer to which—*yes*—has been given by Aharonov & Albert (1981: 369). It all depends upon microcausality: since the *local* observables acting at A must commute with local observables acting at B, the apparent order in which the measurements at A and B take place cannot affect their outcomes. (The question of *nonlocal* observables raises subtle difficulties to which we shall return.) Hence, if Aharonov & Albert are right about this, the enormous theoretical discrepancy between the relativistic and the quantum mechanical theoretical descriptions of EPR-like systems makes no practical difference; indeed, these authors (1981: 369) say that this is how the "...covariance of the probabilities (which is necessary if the theory is to make any sense) and the noncovariance of the state history...manage to peacefully coexist." So long as MC holds up, we can have a relativistic quantum theory, at the price of giving up any hope of describing the reduction process covariantly.

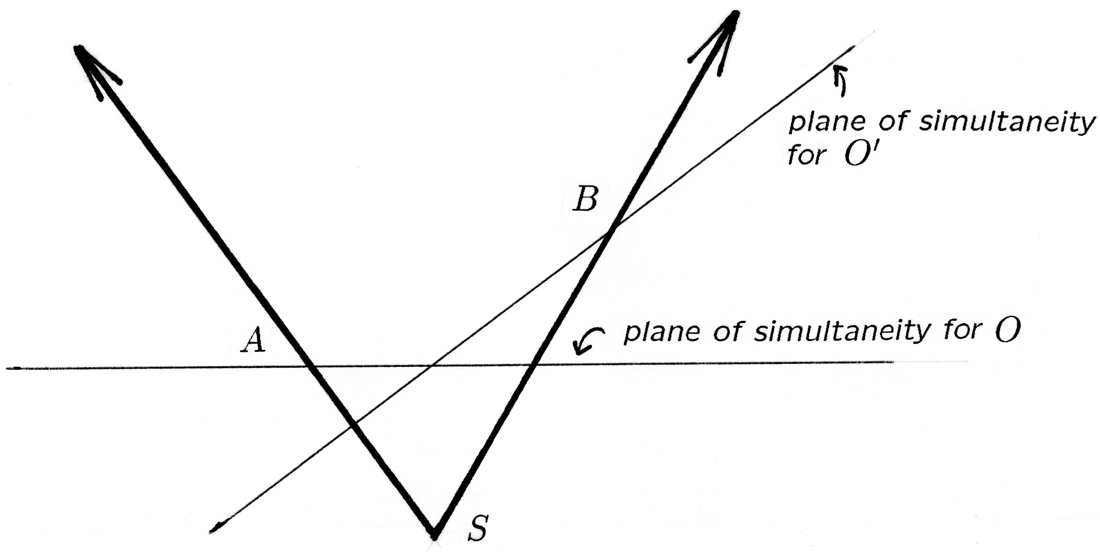

Figure 7.3: O and O' Do Not Agree

It seems to me that there are two problems with this state of affairs. First, this theoretical discrepancy is at least as uncomfortable a thing to live with as a set of books that will not balance, and it would be nice to resolve it somehow. Second, the notion that state reduction is frame-dependent just seems to be wrong. It seems very reasonable to suppose that the location of the amplification event B on the right worldline in Figure 7.1 should be invariant; the reason for this is that this event marks an irreversible change in the properties of the right particle, with an associated increase in entropy, and the spacetime location of such an event should not be observer-dependent. Surely there is a way of matching up our story about reduction with this description of what seems to really happen in such systems.

I will therefore assume, tentatively at least, that in EPR-like experiments such as we have considered here, it is most natural to suppose that the state vector for the total system reduces at a set of invariantly distinguished points (such as A and B in Figure 7.1), or perhaps more generally over some invariant hypersurface which intersects the worldlines of the left and right particles at A and B respectively. This supposition leaves us with two problems:

- Given a specification of the state of the system and the event A, is there some mathematical condition that will allow us to determine the corresponding reduction point B?
- How should we *speak of* the relationship between these two points?

We will deal with these questions in turn.

7.2 The Action Surface

Now, the problem is to find some general criterion or method for identifying the points A and B at which reduction of the system presumably occurs. It turns out that, at least from a perhaps rather naïve viewpoint, there is a surprisingly easy way to do this.

QM tells us that associated with any particle or set of correlated particles there is a wave-packet, which is a superposition of so-called de Broglie waves of the general form[6]

$$\Psi(\mathbf{x}, t) = a_0 e^{i(\mathbf{p}\cdot\mathbf{x} - Ht)/\hbar}. \tag{7.1}$$

Any such waveform is a solution of the time-dependent Schrödinger Equation[7]

$$i\hbar \frac{\partial \Psi(\mathbf{x}, t)}{\partial t} = H\Psi(\mathbf{x}, t). \tag{7.2}$$

The quantity $(\mathbf{p}\cdot\mathbf{x} - Ht)$ represents the action associated with the process, and since action is quantized this must in fact be an *integer*—a fact which in itself is a very strong argument for the discretization of spacetime and momentum-energy coordinates. In the simplest case in which the system moves inertially, so that $(\mathbf{p}, H)$ is just its initial momentum-energy four-vector, the number of action quanta increases steadily with the advance of the system's proper time.

Now, if we really believe in the reduction of the wave packet we must believe that what it amounts to is a sudden "blinking out" of some of the components of the packet. However, it is obviously a mistake to suppose that this "blinking out" is instantaneous in every frame; it must be a covariant process (i.e., a process that has a description as an object in 4-dimensional spacetime). The key to finding the correct 4-dimensional description of this process is that *any component must "blink out" in such a way that all of it disappears at the same point in its cycle.* The condition we will impose, therefore, is *phase invariance*. Suppose the component (7.1) "blinks out" at two spacetime points $(\mathbf{x}, t)$ and $(\mathbf{x}', t')$, with associated points $(\mathbf{p}, H)$ and $(\mathbf{p}', H')$ in momentum-energy space. These must be related such that

$$\mathbf{p}\cdot\mathbf{x} - Ht = \mathbf{p}'\cdot\mathbf{x}' - H't'. \tag{7.3}$$

[6] We let $(\mathbf{x}, t)$ and $(\mathbf{p}, H)$ denote the four-vectors (x, y, z, t) and (p_x, p_y, p_z, H) respectively.
[7] See Dirac (1958: 119–120.)

Therefore any particular component of a wave packet must disappear at spacetime points where the particles of the system have constant *action*.

We can also see that all components that disappear in a particular reduction must do so at the same points, even though these components will not generally be in phase with one another. This is because all the components that disappear together at any *one* spacetime point (which for definiteness could be just A, the point at which the interaction occurs that reduces the system) must disappear everywhere else in such a way as to maintain the same phase relationships.

Therefore, it seems quite reasonable to suppose—at least on the basis of this fairly simplistic model of wave packet reduction—that the points A, B in Figure 7.1 are points of constant *action* for the system. In fact, the reduction will define a hypersurface of constant action in the spacetime of the problem; this hypersurface will represent all the possible points which are candidates to be the actual point B at which the particle's properties become localized. Let us call this hypersurface the *action surface*. In the next section we will see that the relation between the points in this hypersurface can be given a natural interpretation in terms of an invariant simultaneity-like relation. For now, let us consider the implications of this proposal for peaceful coexistence.

Penrose (1986) is one of several authors who have pointed out that state reduction might well be considered to define a "preferred frame" in the sense that it occurs over some particular invariant hypersurface rather than instantaneously in all frames. However, he is reluctant to entertain this possibility because it seems to violate what he calls (1986: 133) the "spirit of relativity". Is this charge really correct?

For once I find myself in defence of peaceful coexistence. First, note that the action hypersurface is not always going to be a hyper*plane* and will therefore not always identify some particular inertial frame as the frame in which the reduction of the system takes place. If it does so, that is essentially an accident of the initial and boundary conditions of the system under study; the particular geometry of the action surface can be expected to depend in a complicated and interesting way upon the particular initial and boundary conditions of the problem, the initial distribution of mass-energy and the initial momenta and positions of the particles involved. There will be nothing lawlike about the geometry of the action surface; the shape it has in a particular setting is just a function of the history of that setting. The action surface hypothesis does not state that there is a special frame of reference or state of motion in which the laws of physics take a special form. The fact that the shape of a given action surface is an invariant is no more a threat to the "spirit of relativity" than is the fact that the particular shape possessed by the surface of the Earth is an invariant.

Of course, it has been long recognized that the phase velocity through space of de Broglie waves is the velocity of wave fronts of constant action (see, for instance, Jammer 1966: 261). But I am not aware that anyone has made the seemingly very obvious inference reported here, which is that if we believe in reduction the disappearance of the wave components—and the very real, thermodynamically irreversible changes of particle state associated with reduction—must occur over these same hypersurfaces of constant action. I suppose that the reason no one has taken this step is because of a reluctance to consider anything that smacks of a lawlike invariant relationship between points at a spacelike separation.

The only author who seems to have been thinking along the same lines as I have outlined here is C. W. Rietdijk (1971, 1985) who has argued (somewhat incoherently, it must be said) for many years that the basic description of quantum phenomena should be in terms of action, not spacetime or momentum-energy. Rietdijk suggests that we regard action as the primary "stuff" of physics and E, t, $\mathbf{x}$, and $\mathbf{p}$ as "derived quantities" (1985: 144); we should *integrate "space-time and energy-momentum jointly into action"*. The real distance between events is not their spacetime or energy-momentum interval but their action difference (their distance in an "internal action metric"). Events which have zero distance in the action metric—such as my amplification points A and B—should be thought of as "contiguous, not really different at all" (1985: 143). Although Rietdijk's mathematics is scarely less sketchy than mine, I think he is certainly on the right track with these ideas, and I will make much use of them in what follows.[8]

7.3 Intrinsic and Extrinsic Simultaneity

We now turn to the second of our problems, which is to find an accurate way of speaking of the relationship between our points A and B, and generally any points on the action surface. The key is to reconsider the possible meanings of the word "simultaneity". What I have to say here is very much in the spirit of Brent Mundy's statement (1986) that our usual usage of this key word contains an essential equivocation, although I will read the equivocation in in a slightly different way than he does.[9]

I would like to make the following suggestion: there are really *two* concepts of simultaneity, which I will call here intrinsic and extrinsic simultaneity. *Extrinsic simultaneity*, which can also be called *coordinate* simultaneity or *metric* simultaneity (Heinlein 1970), denotes equality of measured times; simultaneity with respect to the Einstein simultaneity convention (which Mundy 1986 calls *optical* simultaneity) is an important variety of extrinsic simultaneity. (There are, of course, many other

[8]Rietdijk is also a strong advocate of the so-called retrocausal interpretation of QM. (See also Elitzur 1990.) This is the view that nonlocality can be "explained" by backward-reaching causal influences in spacetime; all we have to do, say the proponents of this theory, is just look at nonlocal systems from the right *four-dimensional* perspective, and the paradoxes all disappear. Without going into detail, I will just note that this seems to be fundamentally mistaken; it is yet another attempt to account for quantum interference phenomena by means of classical trajectories, which is to get things just backwards. In particular, I do not see that these authors would have any way of accounting for interference between *four-dimensional* spacetime paths, such as has recently been observed in two-particle interferometry experiments. See Horne, Shimony & Zeilinger (1990).

[9]Mundy (1986) distinguishes between what he calls *optical* and *causal* simultaneity. Events are optically simultaneous if they have the same time coordinate in some frame of reference, that coordinate having been defined in terms of the Einstein simultaneity convention. Events are causally simultaneous if they cannot be connected by physical influences at finite speed. This is very close to what I mean by intrinsic simultaneity; however, the weakness of Mundy's definition is that "connectible by influences at finite speed" is a noncovariant notion, even though *non-connectibility*, being a topological concept, ought to be an invariant notion.

ways to define a time coordinate.) Two events which are extrinsically simultaneous are said to *happen at the same time*. Einstein taught us that extrinsic simultaneity is a frame-dependent concept; if two events are extrinsically simultaneous in one frame of reference they will not be so in other frames moving relatively to it.

Intrinsic simultaneity, which can also be called *proper* simultaneity, *causal* simultaneity (Mundy 1986), *nonmetric* simultaneity or *topological* simultaneity,[10] is, I propose, an invariant concept. Two events which are intrinsically simultaneous can be said to *happen together*, *happen concurrently*, or *be in joint process*. It is also possible to distinguish a special variety of intrinsic simultaneity which can be called *local* simultaneity; this is the relation between events which Einstein called *coincidence* in the theory of relativity.

One way to introduce the distinction between these two notions of simultaneity is to look at the roots of the word "simultaneous". Generally, it is a poor idea to use etymology as a guide to philosophical analysis, but in this case the history of the word proves to be quite illuminating. The etymon of "simultaneity" is the Latin word *simul*. The *Oxford Latin Dictionary* (1968: 1766) supplies twelve definitions for this interesting word, with many supporting paradigms. One principal meaning for *simul* is, indeed, "At the same time". But it also means "in one and the same action, process, etc., together".[11] The crucial point is that this latter definition *makes no reference to measured time;* it derives its meaning simply from the fact that we have an ability to recognize that two events (generally spatially nearby) can happen together or jointly.

Now I would like to suggest that this other, *non-metric*, sense of "simultaneity" is implicit in our normal way in English of speaking of events and processes.

Let's consider some examples. It is true that we often say such things as "I arrived in Toronto at the same time as she left Paris". But we might also say "The organ played *as* the couple marched down the aisle", or "England slumbered *while* Hitler prepared for war". Although it is customary for us to freely translate such statements into statements about times—"England slumbered during the years that Hitler prepared for war"—the use of such terms as "while" or "as" does not necessarily imply any claim about measured times at all. What they say is that certain events or processes happened *jointly,* as distinguishable parts of some other event or process or set of processes.

A statement that certain events happen together or in joint process can be understood even if one had no clocks, no scale of time at all—even if such things had never been invented. For instance, a small child can clearly understand "Dinner will be served when Daddy comes home", even if she cannot yet read clocks and does not have the slightest understanding of the concept of measured time. There is a rough parallel in spatial language: I might be able to understand what it means to say "We stood side by

[10] I first learned this term from Heinlein (1970); Grünbaum (1973) also uses it, but in a somewhat different and, I think, deliberately ironic way. For Grünbaum, two events are topologically simultaneous if they are outside each other's lightcones. It is easy to see that this is not an equivalence relation (by failure of transitivity) and hence barely qualifies as a simultaneity relation at all.

[11] As Max Jammer has pointed out (lecture at the University of Toronto on October 23, 1989) the Latin word in turn stems from the Sanskit root *sema* which simply means "together" in any respect, spatial or temporal.

side" even if I could not understand "We stood at a very small distance from each other" or "We stood at the same distances from two perpendicular reference lines."

It is crucial that we cannot tell time at all without first having a primitive ability to tell that some events happen in joint process. No scale of time can be defined unless one has the ability to tell that some event of interest occurs as (or while) an accepted nearby clock reads a certain value.[12] A clock, after all, is just a physical system no different in kind from any other except that it exhibits a convenient degree of periodicity; all we are doing, then, when we tell the time of a spatially close event, is noting that the event occurs while a certain nearby clock reading occurs. (Once we can tell the time of nearby events, we can then go on, in the way that Einstein specified, to use local clock readings and light signals to construct a time scale which can define the times of distant events.)

The basis of the view that I am trying to clarify is the recogition that change or process is not itself inherently *metric*—it is not something that comes with a number attached to it. It is *qualitative*. Time and process are not the same thing. As Aristotle said (*Physics* Book B)—and I think he had it right—time is merely a scheme for the measurement or enumeration of motions and changes.[13] And our ability to construct such a scheme depends upon a primitive ability to tell that two spatially nearby events are in process together.[14]

It seems very clear, then, that there are (at least) two significant senses of "simultaneity": the usual meaning, which is "at the same time", which I call *extrinsic simultaneity*, and a more psychologically and epistemologically primitive meaning, which is something like "happening together, in joint process", and which I call *intrinsic simultaneity*.[15]

Now, the kind of intrinsic simultaneity which I have identified so far (*local* simultaneity) is really only meaningful for spatially nearby events.[16] What about distant events—especially spacelike separated events? Is there the slightest indication that it is

[12]See, for instance, Taylor & Wheeler 1966 for a description of how a global time coordinate may be established by exchange of light signals.

[13]Aristotle's careful distinction between time, and motion or process itself, has largely been lost sight of in recent centuries, at the cost of much confusion. In Book B of the *Physics* (Hope translation, 1961) Aristotle says that time is "a measure of movement" (221b); also, time is "a number belonging to movement" (221b); it is "not sheer process but is a numerable aspect of it" (219b); "time is a kind of number" (219b). (In fact, I first heard the notion that "time is a measure of motion" from my friend F. L. Jackson, who thought of it long before either of us had heard of such people as Aristotle.)

[14]The assumption that this is possible is one of the foundation stones of relativity theory. Einstein (Einstein *et al.* 1950: 115):

> We assume the possibility of verifying "simultaneity" for events immediately proximate in space, or—to speak more precisely—for immediate proximity or coincidence in space-time, without giving a definition of this fundamental concept.

[15]This is very close although not identical to what we refer to when we speak of "simultaneous equations" in mathematics. I thank J. R. Brown for this happy observation.

[16]Einstein was well aware that "nearby" is so ill-defined that it beclouds the whole apparently transparent notion of locality. In his paper of 1905 he felt that he had no choice but to brush this puzzle under the carpet, where it has remained ever since. It is worth repeating his remark about this (Einstein 1905, in Einstein *et al.* 1950: 39):

physically meaningful to talk of spatially distant events as happening *as part of the same process?* It turns out that there is.

Consider, again, what apparently happens in a relativistic EPR-like scenario such as I have depicted in Figure 7.1. When a measurement is made on any one particle in the multiparticle system the state vector for the combined system undergoes a global reduction or collapse. The value of certain observables for all the particles will be determined by a measurement on any one of them. The observables did *not*, in general, have definite values before the measurement was made. If we assume that the system was in a definite state (more precisely, a simple product of eigenstates) before the measurement was made, we can derive a Bell Inequality, which is violated by experiment.[17]

Therefore, the *first* thing we can know about the actual state in spacetime of that two-particle system is the appearance of that pair of measurements events. If I get spin up, no further information is required to know that my distant partner got spin down (if she measured the same spin component). There is nothing simpler, or causally or logically prior that I can know about the macroscopic state of the system in spacetime before the measurements were made. There is no machinery, there are no intermediate steps, no hidden or common causes (or if there are, they are *outside* normal space and time). The amplification event triggered by an observation at one end of an EPR apparatus is just an inseparable part of a process by means of which macroscopic conditions in two locations are determined. The whole thing is just one elementary process. Therefore, the two distant events—me getting spin up and my friend getting spin down—can very well be thought of as intrinsically simultaneous in the sense I have indicated here.

In fact, the link between the two events A and B is much stronger than the essentially fortuitous association between nearby events that we have called local simultaneity. The amplification events at either end of an EPR apparatus are *always* coupled in specific ways determined by the relevant conservation laws (such as pertain to charge, linear and angular momentum, mass-energy, etc). The event A *cannot* happen without the event B and vice versa; that is why it is especially meaningful to say they are "in joint process".

Thus, the quantum measurement process defines sets of specially distinguished spacelike related points in spacetime. The points are distinguished in the sense that they are the points at which the values of the observables of the system become fixed.

> We shall not here discuss the inexactitude which lurks in the concept of simultaneity of two events at approximately the same place, which can only be removed by an abstraction.

[17]Bell's Theorem and other recent results such as the Kochen & Specker paradox (see Hughes 1989a) really just confirms that Schrödinger was right when we said (1935, in Wheeler and Zurek 1983: 156):

> "…if I wish to ascribe to the model at each moment a definite (merely not exactly known to me) state, or (which is the same) to *all* determining parts definite (merely not exactly known to me) numerical values, then there is no supposition as to these numerical values *to be imagined* that would not conflict with some portion of quantum theoretical assertions."

I have already given some reason to think that we get into severe inconsistencies if we do not grant that these sets of points are invariant—that is, the same sets of points are distinguished in all frames of reference. Therefore, I propose that we speak of our so-called "reduction" or "amplification" events A and B as being *intrinsically simultaneous*; this is a relationship that holds in all frames even though A and B are, of course, *extrinsically* simultaneous in only one frame of reference. I think that there clearly is, therefore, a frame-independent concept of simultaneity—or, at least, a frame-independent spacelike relation between certain events which is so cognate to our primitive, non-metric notion of simultaneity that it almost certainly deserves to be called a variety of simultaneity.

What we have done, I hope, is found an invariant way of saying that the state function collapse is instantaneous or simultaneous. Of course, I do not claim that this solves or explains the mystery of distant correlations, but I think that it does give us a useful way of speaking of them that is faithful to our *primitive* notions of simultaneity or instantaneity. Certainly some part, at least, of the problem of nonlocality has simply been to find adequate vocabulary.

It is important to emphasize that the intrinsic notion of simultaneity does not pick out the same sets of events as coordinate simultaneity. Recall Figure 7.1: our two "amplification events" A and B are extrinsically simultaneous in only one frame, but the fact that they are in joint process is frame-independent. This lack of extensional equivalence between the two concepts of simultaneity has interesting consequences which I will touch on in a later section.

We are, of course, treading on a minefield here; there are few results of 20th Century physics more jealously guarded by expert opinion than the relativity of simultaneity. Here are typical dicta:

> ... we are to discard our belief in the objective meaning of simultaneity; it was the great achievement of Einstein... that he banished this dogma from our minds... [Weyl 1952: 174]

And:

> ... in special relativity there is no notion of absolute simultaneity; there are no absolute three-dimensional surfaces in spacetime... an observer can still define a notion of which events occur 'at the same time' as a given event... but the notion he gets depends upon his state of motion... The notion that there is absolute simultaneity is a deeply ingrained one. The fact that there is no such notion is one of the most difficult ideas to adjust to in the theory of special relativity. [Wald 1984: 5]

The way around this is to realize that simultaneity does not necessarily mean "at the same *time*", and that in fact this is a derivative sense of the concept. This realization does not actually conflict with the principle of relativity—that there should be a covariant description of all phenomena—but it certainly conflicts with textbook relativity, at least as it is usually (loosely) understood.

To summarize: The most primitive sense of simultaneity is "happening in joint process" and this notion makes no reference to any measured time at which events are presumed to occur. We do not need to understand the notion of measured time to understand the primitive notion of joint process, or happening together, and in fact we need this primitive notion to construct a time scale, since any system of measure is defined in terms of *comparisons* between realities (in this case events) which are not, themselves, numerical. The notion of joint process can easily and naturally be applied to the space-like relation between the "amplification" events in a multi-particle system; in fact, it is stronger than the fairly loose sense of joint process implicit in local simultaneity, since conservation laws guarantee that the two (or perhaps more than two) events at which the local properties of the system become definite cannot happen without each other. Therefore, by paying attention to a distinction not usually made outside the pages of classical dictionaries we can find a very useful way to speak of the invariant, nonlocal relationship between events in an EPR-like system.

To repeat: nothing is transmitted or propagated in an individual measurement in an EPR system; there is just a set of events which happens "at once". What does "at once" mean? It is the event set itself that determines this; the points in spacetime at which these events happen to pop into being are the points which are intrinsically simultaneous, in the sense that they are the loci of events which are "in joint process".

7.4 Intrinsic Simultaneity and Stapp's 'Actual-Events Ontology'

There is a strong parallel between the view I have outlined here and an approach to the metaphysics of quantum physics that Stapp (1989b) has recently called the "actual-events ontology". Stapp states that this is one of the three principal quantum ontologies, the other two being the "pilot-wave ontology" of Bohm and de Broglie and the "many-worlds" ontology of Everett. By "quantum ontologies" Stapp means "conceptions of the constitution of the universe that are compatible with quantum theory." (Stapp 1989b: 269)

The actual-events ontology has a very Aristotelian flavour. It distinguishes between actual events and *potentia* or *possibilia*, and claims that quantum mechanics is a calculus of *possibilia*. (Bohm 1951 expresses a similar view.) Stapp (1989b: 275) says,

> The fundamental process of Nature is taken to be the formation of a sequence of discrete actual events.[18] Each event transforms the potentialities created by the prior event into the potentialities for the next event.

Stapp distinguishes between two kinds of time (1989b: 276):

> ...relativistic quantum theory actually involves two different kinds of time.

[18]Compare Everett's view (Everett 1973: 116): "In the case of observation of one system of a pair of spatially separated, correlated systems, nothing happens to the remote system to make any of its states more 'real' than the rest."

> One of these is the time that is joined to space to form the spacetime continuum of the theory of relativity. This time can be called Einstein time. . . . The second kind of time is connected with the actual changes that occur in connection with quantum jumps. In the Heisenberg representation the Heisenberg state vector is associated with all of spacetime, and the quantum jump, which consists of a change in this state vector, induces changes in expectation values of the field operators over all of spacetime.
>
> The time associated with the quantum jump can be called "process" time. It is associated with *actual* change, whereas Einstein time, in quantum theory, is associated with the evolution of the *potentialities*.

From an Aristotelian point of view, Stapp, while sensing an important distinction, has fallen into an elementary confusion between that which is measured and the descriptive scheme in terms of which the measurements are tallied, for what he calls "process time" is just actual "process" itself. Nevertheless, the distinction he makes is very useful, and, I think, certainly on the right track. If we were to translate my theory of two simultaneities into his terms, we would say that *intrinsic simultaneity is simultaneity with respect to "process time"* while *extrinsic simultaneity is simultaneity with respect to "Einstein time"*.

W. Jones (1989) objects to Stapp's theory on two grounds: the use of two kinds of time is conceptually inelegant, and the theory implies that macroscopic events are well-ordered, "which would be most difficult to reconcile with the experimental evidence for relativity." (Jones 1989: 281) The first objection can possibly be got around by clarifying the distinction between time and process. The second objection raises interesting issues. The term "well-ordered" is due to Jones, and could have been better chosen. Of course, he cannot mean well-ordered in the set-theoretic sense, since all sets can be well-ordered according to a famous theorem of Zermelo. What really concerns Jones is the fact that in any view like mine or Stapp's, all actual events in spacetime are ordered with respect to process time, even though spacelike separate events have no invariant order with respect to Einstein time. But it is not clear that this would conflict with the experimental evidence for relativity (even if it does conflict with some unclear conception of the spirit of relativity). Also, as noted earlier, the particular disposition of "actual events" or "amplification events" in spacetime is determined by initial conditions for the particular system and cosmological boundary conditions; it is a function of the history of the system and hence no threat to the fundamental symmetries of physical theory.

Whether the kind of notions I have sketched here are quite right or not, it is certainly useful to investigate possibilities like these, since a large part of the "radical conceptual renewal" called for by Bell (1986) will almost certainly begin with a careful rethinking of the meaning of *time* and *simultaneity*.

7.5 Simultaneity and the Definition of States

This clarification of the notion of simultaneity, if it is right, has considerable importance for relativistic quantum theory, since it implies that the way we define the state of a system in the relativistic case must be quite different than in the nonrelativistic case. In QM, relativistic or otherwise, a state is defined by a complete set of commuting observables; the possible eigenstates of these observables define a set of basis vectors for the state. To say that these observables commute is to say that they represent measurements which can be carried out on the system *simultaneously*. In nonrelativistic QM the notions of intrinsic and extrinsic simultaneity are *extensionally* equivalent; that is, they pick out the same sets of points in spacetime. Therefore in NRQM (a Galilean-invariant theory) the complete sets of commuting observables are defined over hyperplanes of constant time.

As we have noted, these concepts of simultaneity may not be extensionally equivalent for the relativistic case. What sense of simultaneity is really *implicit* or *intended* in the notion of commuting observables? It surely means that they represent operations which can be carried out together or in joint process; the *times* at which the operations may be performed are irrelevant to their possible simultaneity in this sense. Therefore, if we assume that commutation of observables implies that the observables can be applied to the system in joint process–i.e., that these observables are compatible—then the sense of simultaneity implied by the commutation of these observables must be *intrinsic*; that is, we must conclude that in a properly relativistic QM, complete sets of commuting observables, and therefore states, are fully defined only over hypersurfaces of intrinsic simultaneity, not hypersurfaces of extrinsic simultaneity.

If the action surface hypothesis is correct, therefore, complete sets of commuting observables in spacetime are defined over the action surfaces of the system, not over equal-time hyperplanes in the frames of reference of various observers that may happen to interact with the system. This, of course, is entirely consistent with our conception of how states must collapse, since the reduction of the system ought to amount to a global transition to a new complete set of commuting observables. It offers the hope of a description of the state reduction process that is, indeed, consistent with the "spirit of relativity", even if it does some violence to *textbook* relativity through its admission of a concept of invariant (although not lawlike) simultaneity.

7.6 The Building-Blocks of Spacetime

One of the central premises of classical relativity (i.e., special and general relativity without direct consideration of quantum phenomena) is that spacetime is built up out of elementary coincidences. An observation of a coincidence is taken as epistemologically primitive. As Pathria says (1974: 158):

> . . . all physical measurements, in the last analysis, reduce to a determination of *coincidences* in the spacetime continuum; nothing apart from these coincidences is observable.

The postulate that all observations in the last analysis reduce to observations of spacetime coincidences might almost be taken to be the essence of the classical theory of locality.

Now, I entirely agree with Shimony (1989: 30), who said that "the classical concept of a localized event needs to be broadened", and I think that QM suggests a very natural way to do this. Consider, again, our relativistic EPR scenario. It involves the appearance of a *pair* of spacelike separate, intrinsically simultaneous events A and B. This pair can also be thought of as epistemologically primitive, since the appearance of this pair is the first thing we can know about the classical state of the two particles; in effect, A and B mark the beginnings of two classical worldlines. Therefore, quantum mechanics seems to show that observations of certain kinds of coupled distant events, although less directly accessible to us perceptually, are also a basic modality of observation. There are *local* events and there are *nonlocal* events (i.e., sets of inseparable local events); *both* kinds of events are elementary building blocks of spacetime, and our understanding of spacetime structure up to now must surely have been limited by our consideration of only one of these two kinds of basic events. Pairs (or perhaps even n-tuples) of intrinsically simultaneous spacelike separate events are also elementary building-blocks of spacetime.[19]

I get the following imaginative picture of our quantum universe: classical spacetime and the particles and fields contained in it are built up piece-meal of bricks consisting of sets of correlated point-events, the way an impressionist painting is built up of individual dabs of paint, a photograph is built up of individual amplified photon detections, or the way an image on a video screen is built up of individual scintillations. QM gives us, for reasons that we do not fully understand, an enormous power to calculate the probabilities that these sets of point-events will occur in the various ways that they can occur. Bell's Theorem and other limiting results tell us that if we try to fill in the gaps between the dots with a continuous line, and calculate probabilities as if the particles had a continuous existence in between measurements or observations, we get into a contradiction with experience. This dot-diagram or video-screen picture of physical reality seems to be the absolute minimum that we must accept, whatever further interpretation we choose to put upon it. This leaves unanswered any questions that a realist might have about what lies behind this mysterious popping into spacetime of events. I certainly don't wish to reject such questions as unanswerable or meaningless; I am simply trying to get the description straight of what actually seems to be going on.

Here is Stapp again (1988: 376):

> ... orthodox [quantum] thinking, at least at the informal level, suggests that objective reality is built upon myriads of macroevents which, rising from a sea of micro-level potentialities, create or actualize attributes that weave together to form the fabric of a macroscopic spacetime reality that is describable in terms of the concepts of classical physics.

[19] I have completely glossed over the distinction between fermions and bosons, which is essential in any more thorough development of this theory. The problem of nonlocality is especially acute for fermions, which may, it seems, be *created* in spacelike separate pairs. See Aharonov & Albert 1981: 368–369.

7.7 What Is Causation?

I have noted that there is an obvious problem for this whole way of talking about nonlocality, due to the noninvariance of time order of spacelike related events. Suppose our event A in Figure 7.1 occurs first in a certain frame of reference. Then we are tempted to say that the measurement A *caused* the system to go into a definite state at B. The problem is that event B might occur *before* event A in some other system (because they are spacelike related); causal order is therefore apparently violated. (Penrose (1986) expresses this difficulty engagingly.) The question I will entertain here is simply this: should we really think of A as having *caused* B, or *vice versa*? Or do they stand in some other sort of causal relationship?

It should be clear by now that if the sort of picture I have been painting here is workable, the relation between the correlated events A and B is not causal at all, but is equivalence under membership in a single nonlocal event or process. On this view, the distant correlated events are not related by some kind of action or interaction; they are just aspects of *the same event* popping into spacetime all at once in different locations. Hence there is no question of "action-at-a-distance" between A and B, the kind of thing that worried Einstein so much. The spatially separate events A and B are, in effect, parts of one and the same event—a nonlocal event that is *caused* by the local measurement interaction at one wing of the apparatus.

The major problem that dogs any discussion like this is that we lack an adequate definition of causation in a relativistic context. As Teller says (Cushing & McMullin 1989, p. 212):

> I do not think that anyone has ever made it clear when a sequence of events constitutes a causally connected chain...

It would take me far beyond the scope of this thesis to try to resolve this problem adequately. Nevertheless, we can list (without adequate discussion or proof) a few conditions which would seem to be necessary, although not entirely sufficient, to characterize causation in a quantum relativistic context.

1. As noted above, events can be local or nonlocal. Therefore a causal chain—a sequence of causally related events—can contain local and nonlocal events. For instance, one local measurement can be sufficient to cause the nonlocal appearance of a *set* of amplification events, and these in turn can have local effects on other measuring apparatuses. The No-Signalling Theorem is then the statement that the experimenter who performs the localized measurement cannot control the way the nonlocal result comes out—whether it will be, say, spin up on the left side and down on the right, or *vice versa*. Conversely, if we could signal it would mean that we could somehow exert control over the *entire* nonlocal event, or event-set. Note that even if the process were controllable we would still not say that A causes B; it would still be that the localized interaction near A causes the *pair* of events A, B to appear in a certain way. It is important to distinguish between

the measurement—the *cause*—and the *set* of events $\{A, B\}$—the *effect*. The causal relation is between the local process and the appearance of the pair of causally simultaneous events. The mere fact that a localized event can have an effect at a spacelike separation should not be accounted a violation of causality. Causal anomalies arise only if there are causal loops.

2. Causation is a relation of connectibility between actual events in spacetime. Causal order is therefore invariant or intrinsic; causal order is in terms of the order of events in a causal chain, not in terms of their time ordering in various frames. Time ordering nicely matches causal ordering for timelike causally related events; however, it may not for spacelike causally related events. As noted earlier, this should not be accounted a true causal anomaly, but merely an odd appearance.

3. Causation, whatever it may be, must be a transitive and asymmetric relation.

4. Causation, in some sense that remains to be made precise, must involve a net transfer of *action*. I call this—with apologies to Donald Davidson—the Action Theory of Causation. Then no one event on the same action surface can be the cause of another event on the same action surface—these events are all *causally simultaneous*.

To summarize: I claim that on any reasonable understanding of the terms involved, the amplification events A and B in an EPR-like system ought to be thought of not as causally sequential, but as causally simultaneous. However, this sketchy suggestion obviously needs much more work, and some will certainly find it counterintuitive. One need not accept it in order to accept the second point above, though, namely that causal order, however we define it, is an intrinsic relation between events. This means that the mere fact that different frames of reference will disagree as to the temporal order of spacelike separate events such as our A and B is simply irrelevant to the question of their causal order.

7.8 Are There Causal Paradoxes?

Any theory of spacelike interactions should be judged by how well it accounts for the possibility of closed loop causal anomalies. I will argue here that by this standard the action surface theory, at least in its present sketchy state of development, can be accounted a qualified success.

To begin, I will show how the closed loop "paradox" that we set up, or seemed to set up, back in Section 2.3 can be defused. We can treat this case without having to consider action surfaces or amplification points. Then we will consider a more general case.

Consider Figure 7.4. L and R are two detectors moving away from each other, and S is a source which emits pairs of correlated particles at emission events e_1, e_2, etc. These particles intersect the worldlines of the left and right detectors, respectively, at points A_0, B_0, etc., as shown.

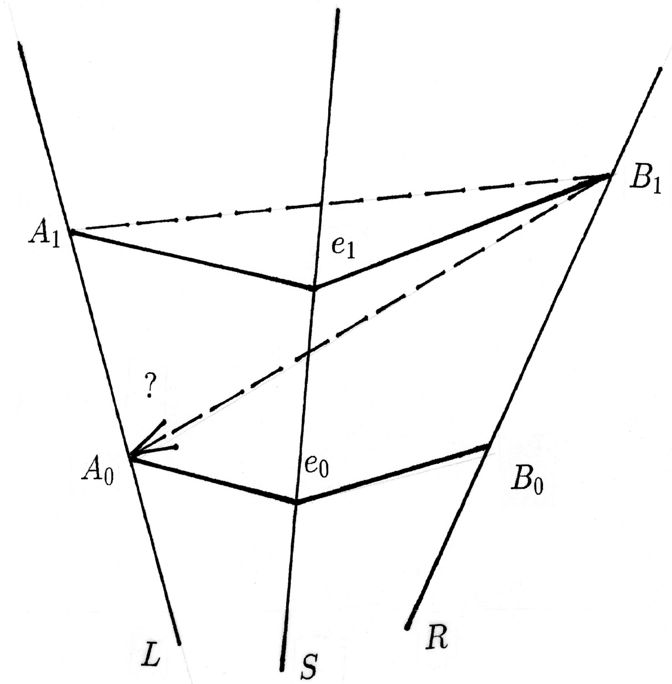

Figure 7.4: Is a Closed Causal Loop Possible Here?

Recall that the basis of this apparent paradox is this: there is a "connection" between observations on the left and right side, based upon the violation of Outcome Independence in experiments of this nature. If we carelessly supposed that this connection is instantaneous in the frame of reference of each detector, then we could perhaps set up a loop such as $A_1 B_1 A_0$, and as I argued earlier the controllability of the interaction is irrelevant to the possibility of setting up an apparent logical contradiction.

The resolution of a "paradox" like this is so trivial that it is *almost* not worth mentioning. If we take care to draw in the worldlines of the source and its daughter particles, we see right away that each event on one worldline can be connected with one and only one event on the other worldline, namely the intersection of the other particle of the pair with the other worldline. In particular, B_1 can only be paired with A_1; it cannot be paired with A_0 because that event belongs to a *different* pair of particles. The right observer might decide to wait until the next particle comes along to send a return signal, but later events on the right worldline can only connect with later events on the left worldline. In either case, no closed loop is possible—so long as later emission events on the source worldline give rise to later detection events on the detector worldlines.

Note that we can turn this argument around to show that in this simple case, in which there is only one source, there is no paradox even if the interaction should happen to be *controllable*; for just as the controllability or otherwise of the interaction is irrelevant to what actually opens up the possibility of a paradox—namely the possibility that right event B_n could be coupled with left event A_{n-i}—the controllability or otherwise of the interaction is irrelevant to what scotches the paradox—namely (i) the fact that each observation event on one side is paired with one and only one event on the other side and (ii) the assumption that the time order of events is the same on all three worldlines.

It has been pointed out to me (by J. R. Brown, in private discussion) that this "paradox" falls apart so easily that it is almost a non-problem. In Brown's words, "There is nothing in entangled states such as the singlet state that would even suggest that measurement on one pair would affect a different pair." But this is precisely the point: *the controllability of the interaction is irrelevant*. Even if the NST were false and one of the experimenters could somehow control which result she got (and thereby affect her partner's distant outcome at will), a simple consideration of the structure of these experiments shows that she could not set up a closed-loop causal paradox; therefore, one of the main concerns about the possibility of signalling seems to have been misplaced.

In summary, in an ordinary EPR experiment, with two detectors in relative motion and one source producing pairs of correlated particles, there is no danger of causal looping, *even if* the results of the nonlocal interactions should be controllable. This is, of course, an argument in favour of peaceful coexistence in one sense, although it turns upon an explicit acceptance of an invariant spacelike relation between events, which might be thought to violate the "spirit of relativity" (as Penrose puts it).

If, now, we allow that there be several sources between the two detector worldlines, the situation gets much more complicated. (I should warn the reader that the rest of this section is the most tentative and conjectural part of this thesis.) To see the sort of problem that can arise, consider Figure 7.5, in which there are two sources shown, each emitting a pair of particles at e_1, e_2.

It looks as if A_2 could be connected with B_2; however, because of the asymmetric placement of the sources, B_2 is earlier than B_1. Therefore it seems that a message could be sent from A_2 to A_1. To see that this is probably not possible we have to make the distinction between amplification events and ordinary measurements by human observers, and try to make some reasonable assumptions about the shape of the action surfaces in the problem.

Recall some of the distinctions we made earlier in this chapter. We noted that the interaction of one of a pair of correlated particles with a localized detector is sufficient to change the global state of the system. Suppose that it is the interaction of the left particle with the left detector that reduces the global state function in Figure 7.6.

We showed that there is good reason to think that the state function reduces over an invariant hypersurface, the hypersurface of constant action through the point A. If this is true, then the state of the right particle reduces at the point B which is the intersection of the action or reduction surface with the worldline of the right particle. We called the events A and B "amplification" events and we noted that the right particle need not interact with a detector *at all* in order for us to be able to make a reasonable

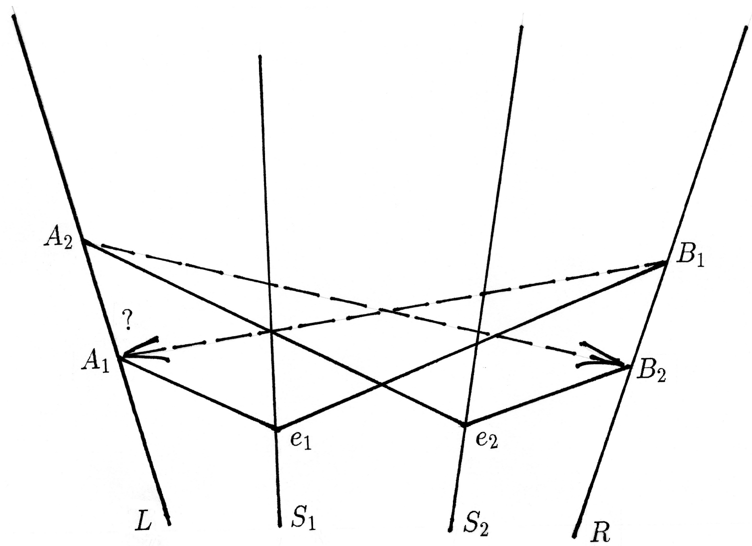

Figure 7.5: Possible Paradox with Multiple Sources

presumption that there is indeed some definite, invariantly distinguished event at which the right particle acquires a definite spin attribute.

If this picture is right, the crucial point to note is that it is only the two amplification events A and B which are "connected" non-classically in the sense that a measurement at A could affect a measurement result at B (were it to be carried out) *or* a measurement at B could affect a measurement result at A. If the right particle does not interact with a detector until some later point B' this later interaction cannot affect the global state of the system, which has already been reduced to a product state by the interaction at A. Therefore *no message, controllable or otherwise, could be sent from B' to A.* The possibility of signalling by means of what we might call "EPR connections" only arises for points on the same action surface.

In Figure 7.5, therefore, merely placing the sources off-centre with respect to the detector is not sufficient to establish the crossing connections shown; the action surface linking A_2 and B_2 would have to intersect the action surface linking A_1 and B_1. Without attempting to justify the claim, I will conjecture that in a "normal" spacetime this never happens: that is, I conjecture that if spacetime is a time-orientable manifold (Wald 1984: 189), then if e_1 and e_2 are events on a timelike worldline such that e_2 is later than e_1, then the action surface through e_2 is everywhere later than the action surface through e_1.[20]

To see that this hopefully plausible assumption does prevent looping, consider Figure 7.7. The spacetime diagram shows the worldlines of two observers who are mov-

[20]"Later" in this context just means later in any coordinate system, since it is an elementary theorem of relativity that time order is invariant for all timelike worldlines.

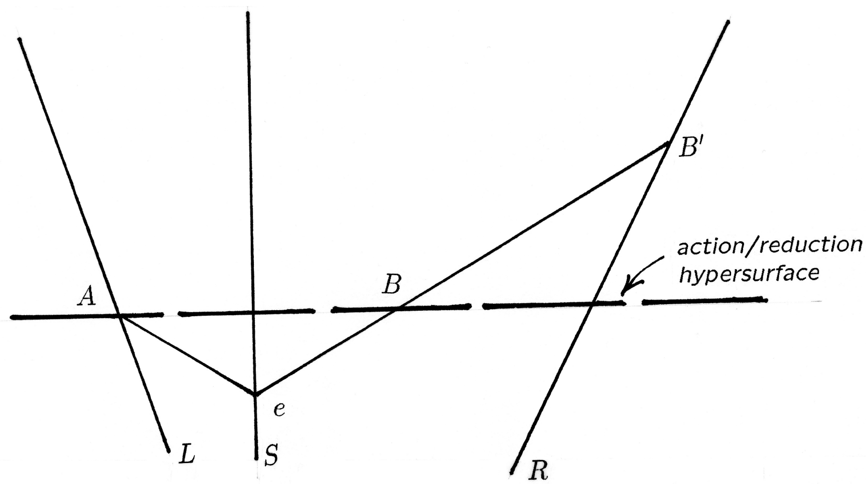

Figure 7.6: No Superluminal Connection Between A and B'

ing away from each other. There is a source or a number of sources of particles between the left and right observer. We don't need to know anything about the disposition of these sources except that the state functions of the particles that they produce reduce over action surfaces Σ_0, Σ_1, and so on.

The observers happen to be placed in such a way as to intersect the action surfaces at amplification points A_0, B_0, and so on. Suppose, for the sake of argument, that the observers can somehow control their local measurement results so that they definitely can get either spin *up* or *down* at will. Then the left observer, for instance, can send a "bit" of information to the right observer by controlling the result at A_0, and the right observer will receive the message at B_0. Now, suppose the right observer wishes to send a reply: the worrisome question is, can she send it in such a way that it will intersect the left observer's worldline at an earlier proper time, for him, than A_0? If so, a closed causal loop can be set up.

The answer is no, and the reason is obvious. The only way the right observer can send a message back to the left is to wait until the next particle intersects her worldline—say at B_1—and then manipulate the result of that measurement in such a way that the left observer gets the result she wants him to get. However, the left observer will invariably intersect the action surface through B_1 at some point on his worldline *later* than the point A_0 at which he sent his message, so long as the action surfaces are neatly layered the way I have shown in the sketch. The only point on the left worldline that B_1 is connnected to is A_1; therefore this is the only point that the right observer can signal to superluminally. So it does not seem to be possible for these observers to set up a closed loop. If this picture is right, one of the principle worries attending the whole notion of nonlocality is removed. (Note, again, that the question of *controllability*

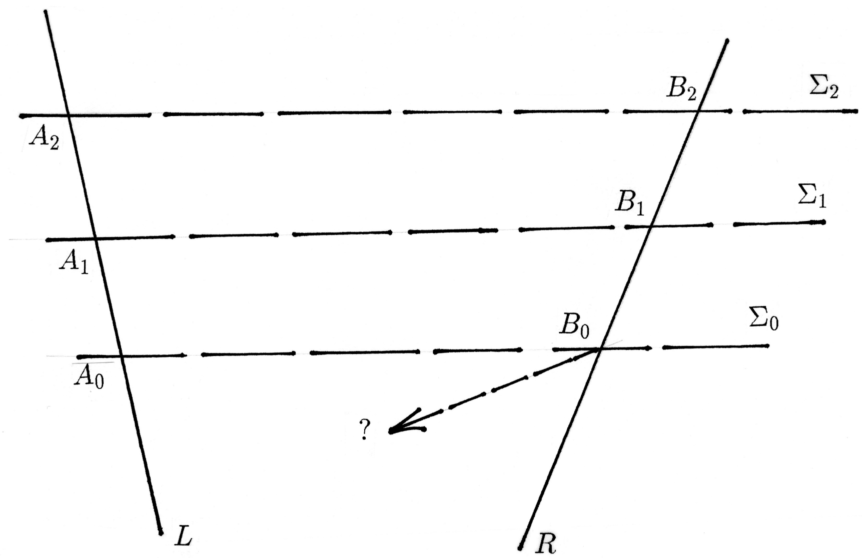

Figure 7.7: Are There Causal Loops?

is not really central; what counts is the topology of the possible connections between spacetime points).

It is, of course, conceivable that there might be folded or intersecting action surfaces. This could happen if there were not a global ordering of time (a possibility which general relativity does seem to permit, at least in principle). Then multiple connections might be possible between points on adjacent worldlines. My guess is that in such a turbulent spacetime there could be no persistent observers to send messages; however, this is wildest speculation. An enormous amount of philosophical and mathematical groundwork must be done before we can even begin to get a reliable picture of the full range of possible relations between spacelike separate amplification points in correlated multi-particle systems. The point is that given a very plausible assumption about the well-ordering of action surfaces, which *probably* holds for all "normal" spacetimes, there is no causal looping with an EPR-like system *even if there are controllable EPR-like interactions*.

Note that this argument does not totally rule out the possibility of closed causal loops *via* superluminal interactions; it merely rules out the possibility of setting up such loops using an EPR system (except possibly in spacetimes of a highly aberrant nature). Suppose that in addition to her measuring apparatus, the right observer has a tachyon gun, which shoots out energy-bearing superluminal particles. Could she not grow impatient with the slowness of the EPR device and try to send a tachyon signal back below Σ_1? (See the dotted arrow at B_0.) I suspect that any such *information-bearing* signal would be ruled out for thermodynamic reasons, since an argument could probably be made that entropy increases drastically across the action surface, which would thus function as a sort of one-way membrane. However, this raises questions which are, obviously, far beyond the scope of this study.

7.9 Microcausality in this Picture

At last we can come to the punchline of the signalling story. Recall our earlier argument that MC could be justified by appeal to the simultaneity of spacelike separate measurements. Obviously, if our distinction between intrinsic and extrinsic simultaneity is cogent, it is *intrinsic* simultaneity that is the relevant sense of simultaneity for this discussion. But, as we have noted, there is good reason to think that not all spacelike related points are intrinsically simultaneous with respect to each other. Hence MC may not hold for *all* operators at a spacelike separation.

MC should hold for observations on points on the same action surface, because these surfaces are the natural surfaces of intrinsic simultaneity in spacetime (Figure 7.8). This means that all observables, local or otherwise, commute on the action surface. For instance, if we measure S_z for a correlated pair of particles, and we get $S_z = +1/2$ at A, we get $S_z = -1/2$ at B, regardless of the time order of A and B in various frames of reference. However, this would not necessarily be the case between A and C. This does *not* imply a violation of spin conservation because the system could have undergone other interactions by the time it evolved to C. We simply cannot state a general commutation relation that would hold between observables acting at points not on the same action surface; the relation between such observables would depend upon the detailed history of the system (the type of interactions it had undergone, etc.).

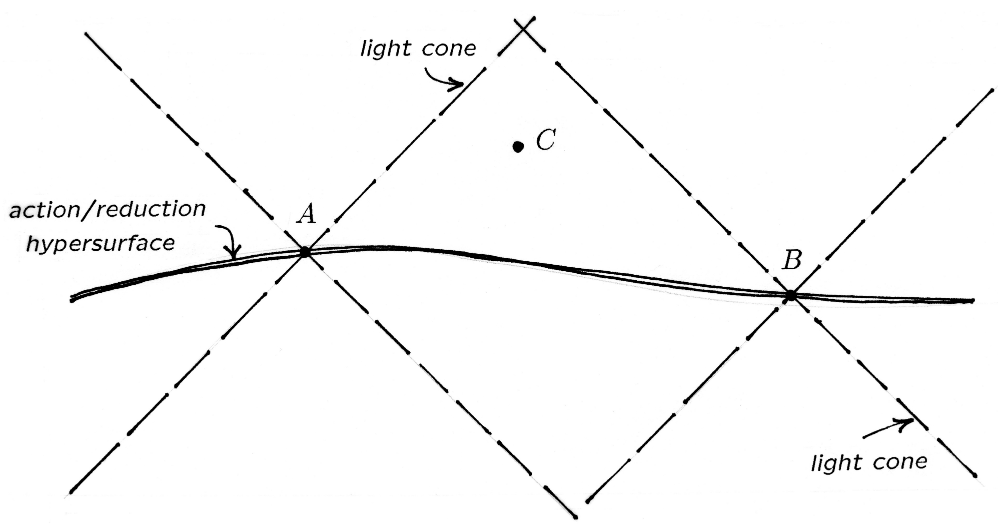

Figure 7.8: Which Points Commute?

We can give a straightforward (although not very rigorous) argument that when A, B are spacelike separate, $[A, B] = 0$ only over the surfaces of constant action. Any commutator in quantum physics has the general form

$$[A, B] = i\hbar C \text{ (units of action)} \tag{7.4}$$

where C is some operator. Now the minimal path connecting any two points not on the same action surface will have non-zero units of action along it. Hence any two points at which A, B occur—and it is really these operations which *define* those points—will have the same action iff the commutator is zero. Proof: if the commutator is zero then the path ABA has zero action; if the path has zero action then the commutator is zero.

It seems very likely, therefore, that MC does *not* hold for spacelike separate points which are not on the same action surface. However, there certainly would be no thermodynamically free transmission between such points; mere local polarizer twiddling will not do it. There would have to be a nonlocal Hamiltonian (recall Shimony's remark) with interaction terms representing the exchange of superluminal particles between such points.

Applying this picture to our problem of signalling, we can summarize matters as follows: if observables are local, then (trivially) there is no signalling. If observables are nonlocal, there is no signalling between points on the same action surface; however, it is not out of the question that there could be superluminal signalling between points not on the same action surface. We have no specific reason to believe that there *is* such a thing; we can only say that, on the view of the spacetime appearance of nonlocality presented here, there is no reason to think that there is *not*. If there were energy-bearing superluminal interactions of some sort then there might be signalling between the points A and C, even if not between A and B. At least, QM cannot offer any general prohibition against the possibility of signalling between points not on the same hypersurface of intrinsic simultaneity; this depends upon detailed consideration of the physics of the particular problem situation. The No-Signalling Theorem therefore offers only qualified and conditional support for the hypothesis of Peaceful Coexistence.

7.10 A Problem and a Prediction

To conclude this overly-long discussion, I am going to "stick my neck out", in the spirit of the Feynman passage at the head of this chapter, and make some predictions.

The thoughtful reader will have noticed that there is one enormous problem with the picture that I have been painting here. *If we admit state reduction only over some invariantly distinguished hypersurface—be it my action surface or any other—we get gross and frame-dependent violations of the conservation of nonlocal quantities such as electrical charge.* The reason is obvious. Consider a frame of reference S in which the amplification event A appears before the corresponding amplification event B, and suppose the pair of particles must obey some conservation law. For the period of time in S between the appearance of A and the appearance of B this conservation law will apparently be violated.

There seem to be two possible responses to this problem. One response (d'Espagnat 1976: 86–91, Aharonov & Albert 1980, 1981) is to essentially give up on the project of finding a covariant description of state reduction, and conclude that the notion of state reduction, and indeed the notion of quantum state itself, simply cannot be made covariant. Aharonov & Albert (1981: 365) conclude:

> The covariance of relativistic quantum theories. . . resides *exclusively* in the experimental probabilities, and not in the underlying quantum states. The states themselves make sense only within a given frame, or, more abstractly, along some given family of parallel spacelike hypersurfaces. . .

The problem with this is that it does not really solve the problem, because there are good reasons to think that the events A and B are not simultaneous in the frame of every observer. If A and B represent the creation of particles, or at least the acquisition of classically definite states by particles, then in order to satisfy conservation laws globally we must allow that these laws may be apparently violated in some frames.

Therefore, another response is to take seriously the possibility that there *are* frame-dependent effects resulting from violations of conservation laws in nonlocal processes, and to explore whether or not these effects might be observable. This may seem very improbable, but it is worth considering.

To be able to make a prediction about such an effect, we need to answer the following question: how can we tentatively identify the hypersurface in our particular spacetime in which state reductions in ordinary Earth-bound quantum mechanics experiments actually happens to occur?[21] Penrose (1986) suggests that this hypersurface might actually be cosmologically determined; this seems eminently reasonable, since if we believe anything like the Big Bang hypothesis, it is reasonable to believe that *all* particles presently existent in the universe (or other particles from which present particles may have happened to decay) have interacted dynamically in the past. Therefore, from the viewpoint of QM, the entire universe may be a coupled system, with one global hypersurface of state reduction. This is, of course, *highly* speculative; nevertheless, it turns out that there is a fairly reasonable and obvious candidate for this hypersurface: the rest frame of the cosmic background radiation.

It is not inconceivable that there might be some sort of experiment which could reveal small violations of conservation laws in the Earth rest frame. These effects would likely be small because our velocity with respect to the cosmic frame, although high by astronomical standards (on the order of 600 km/sec) is nowhere near relativistic. On the other hand, the duration in the Earth frame of the violation would increase if the events A and B were more widely separated in space.

Another possibility might be to try and detect the spatially separated appearance of pairs of particles, perhaps by means of suitably placed detectors such as spark chambers. This is far-fetched, but worth further thought. Perhaps we could actually

[21] The reader will note that I have assiduously avoided all attempts to calculate the geometry of the action surface for any particular system of particles.

see causally simultaneous events “pop” into spacetime. This admittedly improbable experiment could in principle be used to determine the deviation of intrinsic simultaneity from coordinate simultaneity in the Earth frame, as well as check the interpretation offered here. (That is, one could perhaps confirm or disconfirm that the events A and B are, indeed, not necessarily simultaneous in the frame in which they are observed.)

In summary, we can get interesting predictions out of this picture:

- Possible violations of conservation of nonlocal quantities such as charge in Earth’s frame of reference.
- Possible non-simultaneity of amplification events A and B in Earth’s frame; this could be very difficult to detect, if not impossible in principle; on the other hand, detection would be aided by the fact that the lag between A and B would increase with distance.
- More generally, present-day QFT might not retain its validity in reference frames moving at a high velocity with respect to the action frame.

Following a more thorough theoretical analysis, it might be well to search for such effects as these.

Appendix A

Bell's Inequality and the Jarrett Conditions

I reproduce here Shimony's succinct proof (1986: 187) that the Jarrett Conditions together imply a factorizability condition that in turn implies the Bell Inequality.

Recall the terms defined in Section 2.2: p_L and p_R are adjustable detector parameters and x_L and x_R are possible measurement results belonging to the left and right particles respectively. We take as elementary $P_{LR}(x_L, x_R|p_L, p_R)$, the joint probability that measurement results x_L and x_R are obtained together conditional upon parameter settings p_L and p_R. (It may be that from the QM point of view this is the wrong thing to take as elementary, since joint probabilities of this nature can be defined only if the observables are compatible.) Our object is to show that given Parameter and Outcome Independence, this probability factorizes into what might be called localized probabilities:

$$P_{LR}(x_L, x_R|p_L, p_R) = P_L(x_L, \emptyset|p_L, \emptyset)P_R(\emptyset, x_R|\emptyset, p_R). \tag{A.1}$$

We define the localized probabilities as

$$P_{LR}(x_L, \emptyset|p_L, \emptyset) = P_L(x_L|p_L, \emptyset), \tag{A.2}$$

with a corresponding expression for the right particle, and where a measurement result $\emptyset$ simply means that no measurement was made.

For convenience I repeat the definitions of the marginal and conditional probabilities given in Section 2.2,

$$P_L(x_L|p_L, p_R) = \sum_{x_R} P_{LR}(x_L, x_R|p_L, p_R) \tag{A.3}$$

$$P_L(x_L|p_L, p_R, x_R) = \frac{P_{LR}(x_L, x_R|p_L, p_R)}{P_R(x_R|p_L, p_R)}, \tag{A.4}$$

and those of Parameter Independence (PI) and Outcome Independence (OI):

$$P_L(x_L|p_L, p_R) = P_L(x_L|p_L, \emptyset) \quad \text{(PI)} \tag{A.5}$$

$$P_L(x_L|p_L, p_R, x_R) = P_L(x_L|p_L, p_R). \quad \text{(OI)} \tag{A.6}$$

(Corresponding expressions hold, of course, for the right particle.)

Using these definitions, we get

$$
\begin{aligned}
P_{LR}(x_L, x_R|p_L, p_R) &= P_L(x_L|p_L, p_R, x_R) \cdot P_R(x_R|p_L, p_R) \quad \text{using (A.4)} && \text{(A.7)}\\
&= \frac{P_R(x_R|p_L, p_R, x_L)P_L(x_L|p_L, p_R)}{P_R(x_R|p_L, p_R)} \cdot P_R(x_R)|p_L, p_R) \quad \text{using OI} && \text{(A.8)}\\
&= P_L(x_L|p_L, p_R) \cdot P_R(x_R|p_L, p_R) \quad \text{using OI} && \text{(A.9)}\\
&= P_L(x_L|p_L, \emptyset) \cdot P_R(x_R|\emptyset, p_R) \quad \text{using PI} && \text{(A.10)}\\
&= P_L(x_L, \emptyset|p_L, \emptyset) \cdot P_R(\emptyset, x_R|\emptyset, p_R) && \text{(A.11)}
\end{aligned}
$$

This is the factorization condition that entails the Bell Inequality. For a derivation of the B.I. that uses this condition explicitly, see Ballentine (1990: 442) or Selleri (1990: 20–21).

Appendix B

Some Properties of the Singlet State

In this appendix I will briefly summarize the derivation of some of the important properties of the singlet state.

B.1 Definition and Basic Properties

We consider a system of two spin-1/2 particles which are presumed to have decayed from a spin-0 particle. Call the particles L and R (left and right), with associated Hilbert spaces $\mathcal{H}_L$ and $\mathcal{H}_R$ respectively. States of the total two-particle system are elements of the tensor product space

$$\mathcal{H}_{LR} = \mathcal{H}_L \otimes \mathcal{H}_R. \tag{B.1}$$

Spin can be represented by an operator $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$, with

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{B.2}$$

(the Pauli matrices). The operators $\boldsymbol{\sigma}^2$ and σ_z share a common eigenbasis $\{|+\rangle, |-\rangle\}$, the elements of which are kets such that

$$\begin{aligned} \sigma_z|+\rangle &= +|+\rangle \\ \sigma_z|-\rangle &= -|-\rangle. \end{aligned} \tag{B.3}$$

That is, $|+\rangle$ and $|-\rangle$ are eigenvectors of σ_z with eigenvalues ± 1 units of $\hbar/2$. It can easily be shown also that

$$\begin{aligned} \sigma_x|+\rangle &= |-\rangle \\ \sigma_x|-\rangle &= |+\rangle. \end{aligned} \tag{B.4}$$

These kets obey the orthnormalization relations

$$\begin{aligned} \langle +|+\rangle =& \langle -|-\rangle = 1, \\ \langle +|-\rangle =& \langle -|+\rangle = 0. \end{aligned} \tag{B.5}$$

Let $|+\rangle_L$ and $|-\rangle_L$ denote spin states of the left particle, with corresponding notation for the right. The combined system is described by the singlet state

$$\Psi_S = \frac{1}{\sqrt{2}}\big(|+-\rangle - |-+\rangle\big), \tag{B.6}$$

where the notation $|+-\rangle$ is shorthand for

$$|+-\rangle = |+\rangle_L \otimes |-\rangle_R. \tag{B.7}$$

B.2 Correlation Coefficient

Because angular momentum is conserved, we expect measurements of spin in different directions on components of a singlet state to be correlated. We define a correlation coefficient $P(a, b)$ as the average value of the product of spin measurements made on particle L in direction a and R in direction b. Measurements of spin in any direction a are given by the projection of the spin operator on a unit vector in that direction:

$$\text{spin}(a) = \boldsymbol{\sigma} \cdot \hat{\boldsymbol{a}}. \tag{B.8}$$

We can describe the product of results we are interested in as a tensor product operator $\boldsymbol{\sigma_L} \cdot \hat{\boldsymbol{a}} \otimes \boldsymbol{\sigma_R} \cdot \hat{\boldsymbol{b}}$, where $\boldsymbol{\sigma}_L$ and $\boldsymbol{\sigma}_R$ are spin operators in $\mathcal{H}_L$ and $\mathcal{H}_R$ respectively. Then the correlation coefficient we seek is the expectation value of this product:

$$P(a, b) = \langle \Psi_S | \boldsymbol{\sigma_L} \cdot \hat{\boldsymbol{a}} \otimes \boldsymbol{\sigma_R} \cdot \hat{\boldsymbol{b}} | \Psi_S \rangle. \tag{B.9}$$

We set up our apparatus and define our coordinates so that $\hat{\boldsymbol{a}}$ is in the z-direction, and $\hat{\boldsymbol{b}}$ is in the xz-plane. Then we get:

$$\boldsymbol{\sigma}_L \cdot \hat{\boldsymbol{a}} = \sigma_z \equiv \boldsymbol{A} \tag{B.10}$$

$$\boldsymbol{\sigma}_R \cdot \hat{\boldsymbol{b}} = \sigma_z \cos\theta + \sigma_x \sin\theta \equiv \boldsymbol{B}, \tag{B.11}$$

where θ is the angle between $\hat{\boldsymbol{a}}$ and $\hat{\boldsymbol{b}}$. Then we can calculate:

$$\begin{aligned}
P(a, b) &= \frac{1}{2}\big((\langle +- | - \langle -+ |)\boldsymbol{A} \otimes \boldsymbol{B}(|+-\rangle - |-+\rangle)\big) \\
&= \frac{1}{2}\big(\langle + | \boldsymbol{A} | + \rangle \langle - | \boldsymbol{B} | - \rangle - \langle + | \boldsymbol{A} | - \rangle \langle - | \boldsymbol{B} | + \rangle \\
&\quad - \langle - | \boldsymbol{A} | + \rangle \langle + | \boldsymbol{B} | - \rangle + \langle - | \boldsymbol{A} | - \rangle \langle + | \boldsymbol{B} | + \rangle\big) \\
&= \frac{1}{2}\big(\langle - | \boldsymbol{B} | - \rangle - \langle + | \boldsymbol{B} | + \rangle\big) \\
&= \frac{1}{2}\big(-\langle | \sigma_z \cos\theta | - \rangle + \langle - | \sigma_x \sin\theta | - \rangle - \\
&\quad - \langle + | \sigma_z \cos\theta | + \rangle + \langle + | \sigma_x \sin\theta | + \rangle\big) \\
&= \frac{1}{2}\big(-\cos\theta \langle - | - \rangle - \cos\theta \langle + | + \rangle\big) \\
&= -\cos\theta.
\end{aligned} \tag{B.12}$$

B.3 Local Probabilties

By a *local probability* I mean simply the probability that a measurement of a spin component at one branch of the apparatus will yield a given value (say spin *up*). This is given by:

$$\begin{aligned}\left|\langle\sigma_z\cdot\hat{\boldsymbol{b}}=+|\Psi_S\rangle\right|^2 &= \langle\Psi_S|\sigma_z\cdot\hat{\boldsymbol{b}}=+\rangle\langle\sigma_z\cdot\hat{\boldsymbol{b}}=+|\Psi_S\rangle \\ &= \frac{1}{2}\big[(\langle+-|-\langle-+|)(|-+\rangle\langle-+|)(|+-\rangle)-|-+\rangle)\big] \\ &= 1/2, \end{aligned} \tag{B.13}$$

where we have taken $\hat{b}$ along $\hat{z}$. This is a pure number, independent of the measurement procedure carried out on the other particle, a result which in itself effectively constitutes a no-signalling result for the singlet state.

Bibliography

Aharonov, Y. & Albert, D. Z. (1980) "States and observables in relativistic quantum field theories", *Phys. Rev. D* **21** (12), 1316–1325.

——— (1981) "Can we make sense out of the measurement process in relativistic quantum mechanics?", *Phys. Rev. D* **24** (2), 359–370.

Ardavan, H. (1984a) "A speed-of-light barrier in classical electrodynamics", *Phys. Rev. D* **29**(2), 207–215.

——— (1984b) "A singularity arising from the coherent generation of gravitational waves by electromagnetic waves", in Bonnor, W. B., Islam, J. N., & MacCallum, M. (eds.) *Classical General Relativity* Cambridge University Press, Cambridge, 5–14.

Aspect, A., Grangier, P., & Roger, G. (1981) "Experimental tests of realistic local theories via Bell's Theorem", *Phys. Rev. Lett.* **47**(1), 460–463.

——— (1982) "Experimental realization of Einstein-Podolsky-Rosen-Bohm *Gedankenexperiment*: A new violation of Bell's Inequalities", *Phys. Rev. Lett.* **49**(2), 91–94.

Aspect, A., Dalibard, J., & Roger, G. (1982) "Experimental test of Bell's Inequalities using time-varying analysers", *Phys. Rev. Lett.* **49**(25), 1804–1807.

Bacry, Henri (1988) *Localizability and Space in Quantum Physics*, Springer-Verlag, Berlin.

Ballentine, Leslie E. (1990) "Limitations of the Projection Postulate", *Foundations of Physics* **20**(11), 1329–1343.

——— (1990) *Quantum Mechanics*, Prentice Hall, Englewood Cliffs, New Jersey.

Ballentine, L. E., & Jarrett, J. P. (1987) "Bell's Theorem: Does quantum mechanics contradict relativity?", *Am. J. Phys.* **55**(8), 696–701.

Bell, J. S. (1964) "On the Einstein-Podolsky-Rosen Paradox", *Physics* **1**, 195–200; reprinted in Bell (1987b), 14–21.

——— (1966) "On the problem of hidden variables in quantum mechanics", *Rev. Mod. Phys.* **38**, 447–452; reprinted in Bell (1987b), 1–13.

——— (1986) "Introductory Remarks", *Physics Reports* **137**(1), 7–9.

——— (1987a) "Are there quantum jumps?" in Kilmister (ed.) (1987), 41–52.

——— (1987b) *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, Cambridge.

——— (1988) Interview by C. Mann and R. Crease *Omni*, May, pp. 84+.

——— (1990) "Against 'measurement' ", *Physics World* **3**(8), 33–40.

Bernstein, Jeremy (1991) *Quantum Profiles*, Princeton University Press, Princeton.

Bilaniuk, O. & Sudarshan, E. (1969) "Particles beyond the light barrier", *Physics Today*, May.

Bjorken, J. D., & Drell, S. D. (1965) *Relativistic Quantum Fields,* McGraw-Hill, New York.

Blum, Karl (1981) *Density Matrix Theory and Applications*, Plenum Press, New York and London.

Bogolubov, N. N., Logunov, A. A., & Todorov, I. T. (1975) *Introduction to Axiomatic Quantum Field Theory*, Benjamin/Cummings, Reading.

Bohm, David (1951) *Quantum Theory*, Prentice-Hall, Englewood Cliffs.

——— (1980) *Wholeness and the Implicate Order*, Routledge & Kegan Paul, London.

Bohr, Niels (1958) *Atomic Physics and Human Knowledge*, Wiley, New York.

——— (1963) *Essays 1958–1962 on Atomic Physics and Human Knowledge*, New York.

Braginsky, V. B., Vorontsov, Y. I., & Thorne, K. S. (1980), "Quantum nondemolition measurements", *Science* **209**, 547–557, reprinted in Wheeler & Zurek (eds.) (1983), 749–768.

Brown, J. R. (1991) *The Laboratory of the Mind*, Routledge, London.

Bussey, P. J. (1982) " 'Super-luminal communication' in Einstein-Podolsky-Rosen Experiments", *Physics Letters*, **90A** Number 1,2, 9–12.

Chester, Marvin (1987) *A Primer of Quantum Mechanics*, John Wiley and Sons, New York.

Cohen-Tannoudji, C., Diu, B., & Laloë, F. (1977) *Quantum Mechanics*, Vols. I & II, John Wiley & Sons, New York.

Copi, Irving (1982) *Introduction to Logic*, Sixth Edition, Macmillan, New York.

Cushing, James T. (1989) "A background essay", in Cushing & McMullin (eds.) (1989), 1–24.

Cushing, James T., & McMullin, Ernan (eds.) (1989) *Philosophical Consequences of Quantum Theory: Reflections on Bell's Theorem*, University of Notre Dame Press, Notre Dame.

Davies, P. C. W., & Brown, J. (1986) *The Ghost in the Atom*, Cambridge University Press, Cambridge.

d'Espagnat, B. (1976) *Conceptual Foundations of Quantum Mechanics*, Second Edition, Addison-Wesley Pub. Co., New York.

——— (1984) "Nonseparability and the tentative tDescriptions of reality", *Physics Reports* **110**(4), 201–264. Reprinted in Schommers (ed.) (1989).

——— (1988) "Special Preface" to reprint of d'Espagnat (1976).

DeWitt, B. S. & Graham, N. (eds.) (1973) *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton University Press, Princeton.

Dirac, P. A. M. (1958) *The Principles of Quantum Mechanics*, Fourth Edition (Revised 1967), The Clarendon Press, Oxford.

Eberhard, P. H. (1978), "Bell's Theorem and the different concepts of locality", *Nuovo Cimento* **46B**, 11(2), 392–419.

——— (1989a) "The EPR Paradox. Roots and ramifications", in Schommers (ed.) (1989), 49–88.

——— (1989b) "A realistic model for quantum theory With a locality property", in Schommers (1989), 169–215.

Eberhard, P. H., & Ross, R. R. (1989) "Quantum Field Theory cannot provide faster-than-light communication", *Foundations of Physics Letters* **2**(2), 127–149.

Einstein, A. (1905) "Zur Elektrodynamik bewegter Körper", *Annalen der Physik* **17**.

——— (1919) "What is the Theory of Relativity?" tr. by V. Bargman in Einstein (1954).

——— (1954) *Ideas and Opinions*, Crown, New York.

Einstein, A. *et al.* (1950) *The Principle of Relativity*, tr. by W. Perrett & G. B. Jeffrey, Dover, New York.

Elitzur (1990) "Neglected thermodynamic considerations", *Foundations of Physics Letters* **3**(6).

Everett, Hugh (1973) "The Theory of the universal wave function", in DeWitt & Graham (eds.) (1973), 3–140.

Feynman, R. P., Leighton, R., & Sands, M. (1965) *The Feynman Lectures in Physics, Vol. III: Quantum Mechanics*, Addison-Wesley, Reading, Mass.

Fine, Arthur (1986) *The Shaky Game: Einstein, Realism and the Quantum Theory*, University of Chicago Press, Chicago and London.

——— (1989) "Do correlations need to be explained?" in Cushing & McMullin (eds.) (1989), 175–194.

Franson, J. D. (1989) "Bell Inequality for position and time", *Phys. Rev. Lett.* **62**(19), 2205–2208.

Ghirardi, G. C. (1988) "Some critical considerations on the present epistemological and scientific debate on quantum mechanics", in Tarozzi & van der Merwe (1988), 89–103.

Ghirardi, G. C., Grasssi, R., & Pearle, P. (1990) "Relativistic dynamical reduction rodels: General framework and examples", *Foundations of Physics* **20**(11), 1271.

Ghirardi, G. C., Rimini, A., & Weber, T., (1980) "A general argument against superluminal transmission through the quantum mechanical measurement process", *Lettere al Nuovo Cimento*, **27**(10), 293–298.

Ghirardi, G. C., & Weber, T. (1983) "Quantum Mechanics and faster-than-light communication: Methodological considerations", *Nuovo Cimento* **78B**(1), 9–20.

Grünbaum, A. (1973) *Philosophical Problems of Space and Time*, 2nd Edition, Reidel, Dordrecht.

Gudder, Stanley P. (1977), "Four approaches to axiomatic quantum mechanics", in Price & Chissick, 247–276.

Hannabus, K. C. (1988) *An Introduction to Quantum Mechanics*, Mathematical Institute, University of Oxford.

Heinlein, Robert A. (1970) *Orphans of the Sky,* Berkeley (reprint), New York.

——— (1978) *Time for the Stars,* Ballantine (reprint), New York.

Herbert, Nick (1982) "FLASH—A superluminal communicator based upon a new kind of quantum measurement", *Foundations of Physics* **12**(12), 1171–1179.

——— (1988) *Faster Than Light: Superluminal Loopholes in Physics*, New American Library, New York.

Horne, Michael, Shimony, Abner, & Zeilinger, Anton (1990), "Two-particle interferometry", *Nature* **347**, 4 October, 429–430.

Howard, Don (1985) "Einstein on locality and separability", *Stud. Hist. Phil. Sci.*, **16**(3), 171–201.

Hughes, R. I. G. (1989a) *The Structure and Interpretation of Quantum Mechanics*, Harvard University Press, Cambridge, Massachusetts.

——— (1989b) "Bell's Theorem, ideology, and structural explanation", in Cushing & McMullin (eds.) (1989),195–207.

Jammer, Max (1966) *The Conceptual Development of Quantum Mechanics*, McGraw-Hill, New York.

Jarrett, Jon P. (1984), "On the physical significance of the locality conditions in the Bell arguments", *Noûs* **18**, 569–589.

——— (1989) "Bell's Theorem: A guide to the implications", in Cushing & McMullin (1989), 60–79.

Jauch, J. M., & Rohrlich, F. (1955) *The Theory of Positrons and Electrons*, Addison-Wesley, Reading, Mass.

Jones, M. R., Clifton, R. K., & Masters, A. (1990) "Completeness, locality and superluminal signalling", Abstract in *APA Proceedings* **64**(4), 88.

Jones, William B. (1989) "Henry P. Stapp on quantum theory and reality", in Kefatos (ed.) (1989), 279–281.

Jordan, T. F. (1983) "Quantum correlations do not transmit signals", *Physics Letters* **94A**, Number 6,7, 264.

Kafatos, Menas (ed.) (1989) *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, Kluwer Academic Publishers, Dordrecht.

Kaufmann, W. J. (1977) *The Cosmic Frontiers of General Relativity,* Little, Brown, & Co., Boston and Toronto.

Kekez, D., Ljubicic, A., & Logan, B. A. (1990) "An upper limit to violations of the Pauli exclusion principle", *Nature* **348**, 15 November 1990, 224.

Kilmister, C. W. (ed.) (1987) *Schrödinger: Centenary Celebration of a Polymath*, Cambridge University Press, Cambridge.

Knuth, Donald E. (1986) *The TEXbook*, Addison-Wesley, Reading.

Lakatos, Imre (1976) *Proofs and Refutations: The Logic of Mathematical Discovery*, Cambridge, Cambridge University Press.

Landau, L. D., & Lifshitz, E. M. (1977), *Quantum Mechanics*, (Third Edition), Pergamon.

Lee, T. D. (1981) *Particle Physics and Introduction to Field Theory*, Harwood, New York.

Mandl, F. (1959) *Introduction to Quantum Field Theory*, Interscience Publishers, New York.

McMullin, E. (1989) “The explanation of distance action: Historical notes”, in Cushing & McMullin (eds.) (1989), 272–302.

Misner, C. W., Thorne, K. S., & Wheeler, J. A. (1973) *Gravitation*, W. H. Freeman, San Francisco.

Mundy, B. (1986a) “The physical content of Minkowski geometry” *Brit. J. Phil. Sci.* **37**, 25–54.

——— (1986b) “Special Relativity and quantum measurement” *Brit. J. Phil. Sci.* **37**, 207–212.

Oxford Latin Dictionary, A. Souter *et al.* (eds.) (1968), Clarendon Press, Oxford.

Page, D. N. (1982) “The Einstein-Podolsky-Rosen Physical reality is completely described by quantum mechanics”, *Physics Letters* **91A**(2), 57–60.

Pathria, R. K. (1974) *The Theory of Relativity*, Pergamon, Oxford.

Pauli, W. (1940) “The connection between spin and statistics”, *Physical Review* **58** 716–722; reprinted in Schwinger, J. (ed.) (1958) *Selected Papers on Quantum Electrodynamics*, Dover, New York, 372–278.

Penrose, R. (1986) “Gravity and state vector reduction”, in Penrose & Isham (1986), 129–146.

Penrose, R. (1989) *The Emperor's New Mind*, Vintage, London.

Penrose, R. & Isham, C. J. (eds.) (1986) *Quantum Concepts in Space and Time*, Clarendon Press, Oxford.

Pitowski, Itamar (1989) “From George Boole to John Bell — The origins of Bell's Inequality”, in Kafatos (ed.) (1989), 37–49.

Polchinski, J. (1991) “Weinberg's nonlinear quantum mechanics and the Einstein-Podolsky-Rosen Paradox”, *Phys. Rev. Lett.* **66**(4), 397–400.

Price, William C., & Chissick, S. S. (eds.) (1977), *The Uncertainty Principle and Foundations of Quantum Mechanics: A Fifty Years Survey*, John Wiley and Sons, London & New York.

Prigogine, Ilya (1980) *From Being to Becoming*, Freeman, San Francisco.

Recami, E. (1977), “A time operator and the time-energy uncertainty relation”, in Price & Chissick, 21–28.

——— (1987) *Foundations of Physics* **17**(3), 239.

Redhead, Michael (1983), “Nonlocality and Peaceful Coexistence”, in R. Swinburne (ed.) (1983), 151–189.

——— (1987), *Incompleteness, Nonlocality and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*, Clarendon Press, Oxford.

Rietdijk, C. W. (1971) *On Waves, Particles and Hidden Variables: A New Approach*, Van Gorcum & Co., Assen.

——— (1985) “On nonlocal influences”, in Tarozzi and van der Merwe (eds.) (1985), 129–151.

Roman, Paul (1969), *Introduction to Quantum Field Theory*, John Wiley and Sons, New York.

Sánchez-Gómez, J. L. (1988) “Comments on relativistic covariance and causality in EPR problems”, in van der Merwe *et al.* (eds.) (1988), 393–399.

Schilpp, Paul (ed.) (1969) *Albert Einstein, Philosopher-Scientist,* Open Court, La Salle. (Third edition).

Schommers, W. (ed.) (1989) *Quantum Theory and Pictures of Reality*, Springer-Verlag, Berlin.

——— (1989) "Space-time and quantum phenomena", in Schommers (ed.) (1989), 217–277.

Schrödinger, E. (1984) *Collected Papers*, Austrian Academy of Science, Vienna; cited in Bacry (1988).

Schweber, S. S. (1961) *An Introduction to Relativistic Quantum Field Theory,* Row, Peterson and Co., Evanston.

Selleri, Franco (1990) *Quantum Paradoxes and Physical Reality,* Kluwer Academic Publishers, Dordrecht.

Shimony, A. (1978), "Metaphysical problems in the foundations of quantum mechanics", *International Philosophical Quarterly* **18**(1), 3–17.

——— (1983) "Controllable and uncontrollable non-locality", in S. Kamefuchi *et al.* (eds.), *Proc. Int. Symp. Foundations of Quantum Mechanics*, Physical Society of Japan, Tokyo, pp. 225–230.

——— (1984) "Contextual hidden variables theories and Bell's Inequalities", *Brit. J. Phil. Sci.* **35**, 25–45.

——— (1986) "Events and processes in the quantum world", in Penrose & Isham (1986), 182–203.

——— (1989) "Search for a worldview which can accomodate our knowledge of microphysics", in Cushing & McMullin (eds.) (1989), 25–37.

Stapp, H. P. (1988) "Einstein locality, EPR locality, and the significance for science of the nonlocal character of quantum theory", in van der Merwe *et al.* (eds.) (1988), 367–378.

——— (1989a) "Comments on 'Quantum theory does not require action at a distance' ", *Foundations of Physics Letters* **2**(1), 9–13.

——— (1989b) "Quantum ontologies", in Kefatos (ed.) (1989), 269–277.

Streater, R. F. (1987) "Why axiomatize quantum field theory?" in Brown & Harré (eds.) (1988), 137–148.

Streater, R. F., & Wightman, A. S. 1964, *PCT, Spin and Statistics and All That*, Benjamin, New York.

Tarozzi, G., & van der Merwe, A. (eds.) (1985) *Open Questions in Quantum Physics*, D. Reidel Publishing Co., Dordrecht.

Tarozzi, G., & van der Merwe, A. (eds.) (1988) *The Nature of Quantum Paradoxes*, Kluwer Academic Publishers, Dordrecht.

Taylor, E. F., & Wheeler, J. A. (1966) *Spacetime Physics*, Freeman, San Francisco.

Teller, Paul (1989) "Relativity, relational holism, and the Bell Inequalities", in Cushing & McMullin (1989), 208–223.

Troudet, T. (1985) "Aharonov-Bohm Effect versus causality?" *Physics Letters* **111A**(6) 274–276.

van der Merwe, A., Selleri, F., & Tarozzi, G. (eds.) (1988) *Microphysical Reality and Quantum Formalism, Vol. II*, Kluwer Academic Publishers, Dordrecht.

van Fraassen, Bas C. (1989) "The Charybidis of Realism: Epistemological implications of Bell's Inequality", in Cushing & McMullin (1989), 97–113.

van Kampen, N. G. (1984) “Can the Aharonov-Bohm Effect transmit signals faster than light?” *Physics Letters* **106A**(1,2) 5–6.

Wald, Robert H. (1984) *General Relativity,* University of Chicago Press, Chicago and London.

Weyl, H. (1952) *Space-Time-Matter*, Dover, New York. (Translation by H. L. Brose of 4th edition of *Raum. Zeit. Materie.* Springer, Berlin, 1921.)

Wheeler, J. A. (1983) “Law without law”, in Wheeler & Zurek (eds.) (1983), 182–213.

Wheeler, J. A., & Taylor, E. (1963) *Spacetime Physics*, Freeman, San Francisco.

Wheeler, J. A., & Zurek, W. H. (eds.) (1983) *Quantum Theory and Measurement*, Princeton University Press, Princeton.

Wigner, Eugene (1963) “The problem of measurement”, *Am. J. Phys.* **31**, 6–15.

Woodward, P. M. (1953) *Probability and Information Theory, With Applications to Radar*, Pergamon Press, London.