December 1995 LBL-38076

Dealing with quantum weirdness: Holism and related issues. *

Ph.D Thesis

Andrew Richard Elby

Theoretical Physics Group
Lawrence Berkeley National Laboratory
University of California
Berkeley, California 94720

*This work was supported by the Director, Office of Energy Research, Office of High Energy and Nuclear Physics, Division of High Energy Physics of the U.S. Department of Energy under Contract DE-AC03-76SF00098.

MASTER

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Recycled Paper

Abstract

Dealing with quantum weirdness: Holism and related issues

by

Andrew Richard Elby

Doctor of Philosophy in Interdisciplinary Studies: Philosophy of Physics

University of California at Berkeley

Professor Geoffrey Chew, Chair

I discuss a variety of issues in the interpretation of quantum mechanics. All of these explorations point toward the same conclusion, that some systems are holistically connected. In other words, some composite systems possess properties that cannot, even in principle, be reduced to (or "built up" from) the properties of its subsystems. This, I argue, is a central metaphysical lesson of quantum theory, a lesson that will pertain even if quantum mechanics eventually gets replaced by a superior theory.

After outlining this dissertation in chapter 1, I jump into issues of nonlocality in chapter 2. There, I establish a new, probabilistic framework in which to formulate "algebraic" (perfect correlations) nonlocality proofs. Working within that framework, I rule out hidden-variable theories that approximately reproduce the perfect correlations of

quantum mechanics, as well as theories that obey locality conditions weaker than those needed to derive Bell's inequality.

In chapter 3, I discuss Superconducting Quantum Interference Devices (SQUIDs). *Contra* Leggett, I show that SQUID experiments cannot rule out Macrorealism. What they *can* rule out is non-invasive measurability, the assumption that it's possible (in principle) to measure a system with arbitrarily small disturbance to its future dynamics. Failure of non-invasive measurability is best explained as resulting from a holistic connection between measuring device and measured system.

Chapter 4 looks at the interpretational issues surrounding decoherence, the dissipative interaction between a system and its environment. Decoherence alone neither constitutes nor points to a specific interpretation of the quantum formalism. It can, however, help "modal" interpretations pick out the desired "preferred" basis. After raising some potentially fatal objections to the modal interpretation, I show in detail how decoherence comes to the rescue. Modal interpretations explicitly incorporate holism.

Finally, in chapter 5, I explore what varieties of causation can and cannot "explain" the EPR correlations. Any purported causal explanation of EPR within the context of relativistic quantum theory must renounce the "generative" causal intuition that causes bring about their effects. I explore the philosophical ramifications of this result, concluding that instead of relying upon "watered down" causal explanations, we should instead develop new, holistic explanatory frameworks.

Dedication

This dissertation wouldn't exist but for the stimulation and inspiration I've received from colleagues all over the world. So, if you don't like this thesis, don't blame me. Blame them. Heh heh.

I'd particularly like to thank Guido Bacciagaluppi, Harvey Brown, Jeff Bub, Jeremy Butterfield, Geoffrey Chew, Rob Clifton, Sara Foster, Richard Healey, Meir Hemmo, Martin Jones, Michael Redhead, Allen Stairs, and Paul Teller.

I dedicate this dissertation to Diana Perry, who is far more intellectually interesting than what follows.

Contents

Preface

Hi, Andy here. So, you actually want to read (parts of) this dissertation. I thought it would never happen, but hey, I'm happy to be proven wrong.

If you're an expert on any of these topics, you might not want to read the corresponding chapter of this dissertation, which spends a fair amount of time establishing context and explaining the central issues to interested philosophers and physicists. Instead, you might consider looking up my published papers on these topics. (See the references at the end of this dissertation for a partial list.) I've tried to write these thesis chapters in such a way that they could serve to introduce someone to the field, in the style of a good review paper.

CHAPTER 1: OVERVIEW

Does quantum mechanics force us, or at least invite us, to revise our metaphysical views? And if so, how radically? In this dissertation, I'll argue that we should abandon a causal explanatory framework in favor of a holistic one. By "holism," I mean that a composite system can possess properties that are not reducible to, and cannot be "built up" from, the properties of its parts, even in principle.

This argument is hard to make, because many examples of "quantum weirdness" can be explained nonholistically. Indeed, none of my chapters directly argue for a holistic world view. Instead, my strategy is to show, in a variety of quantum contexts that a holistic outlook is one of only two or three viable alternatives. So, none of my chapters on its own favors holism. But since all the chapters are about different issues in philosophy of quantum mechanics, and since they all establish that a holistic outlook is one of only a few viable alternatives, the dissertation as whole manages to argue--almost by "brute force"--that holism deserves a closer look.

For instance, consider the nonlocality explored in chapter 2. There, I establish a new, probabilistic framework in which to formulate "algebraic" (perfect correlations) nonlocality proofs. Working within that framework, I rule out hidden-variable theories that approximately reproduce the perfect correlations of quantum mechanics, as well as theories that obey locality conditions weaker than those needed to derive Bell's inequality. We can interpret this irreducible nonlocality as resulting from

(A) a superluminal causal influence, or from direct action at a distance; or

(B) a holistic connection between the two wings of the experiment.

Explanation (A), though unsettling and problematic, stays within a "classical causal" framework, in which individual systems all "have" separate (nonholistic) properties, and causal interactions are what bring about changes in these properties. Chapter 2 can't pick out which metaphysical outlook--holism or causal nonlocality--is the best choice. It can merely present you with those options.

In chapter 3, I discuss Superconducting Quantum Interference Devices (SQUIDs). *Contra* Leggett, I show that SQUID experiments cannot rule out Macrorealism. What they *can* rule out is non-invasive measurability, the assumption that it's possible (in principle) to measure a system with arbitrarily small disturbance to its future dynamics. Failure of non-invasive measurability could be explained in terms of

(A) a weird kind of causal interaction, the severity of which cannot be made very small even in principle; or

(B) a holistic entanglement between the measuring device and measured system.

Again, I can't firmly establish that we should choose (B) over (A). But it's intriguing that the causal alternatives in chapters 2 and 3--choice "A" in both couplets--are problematic *for different reasons.*

Chapter 4 looks at the interpretational issues surrounding decoherence, the dissipative interaction between a system and its environment. Decoherence alone neither constitutes nor points to a specific interpretation of the quantum formalism. It can, however, help so-called "modal" interpretations pick out the desired "preferred" basis. After raising some potentially fatal objections to the modal interpretation, I show in detail how decoherence comes to the rescue, making modal interpretations one of the

few viable classes of interpretations currently out there. Modal interpretations explicitly incorporate holism. The fact that an aggressively holistic interpretation can "harness" decoherence in order to explain our familiar classical reality makes it plausible that holism is compatible with everyday experience. Once again, I can't argue that the holistic story here is better than (say, Bohm's) causal story. But the holistic modal interpretation, with some further work, may be just as viable. (And besides, Bohm's theory may well incorporate holism, too.)

Finally, in chapter 5, I explore what varieties of causation can and cannot "explain" the EPR correlations. Any purported causal explanation of EPR must renounce the "generative" causal intuition that causes bring about their effects. I explore the philosophical ramifications of this result, concluding that instead of relying upon "watered down" causal explanations, we can either

(A) accept one of these eviscerated causal explanations, or

(B) develop new, perhaps holistic explanatory frameworks.

As you can see, no one of my chapters should convince you that holism is the way to go. But taken together, they present a strong case.

CHAPTER 2: PERFECT-CORRELATIONS NONLOCALITY PROOFS

Section 2.1: Introduction

With the exception of Selleri and his "enhancement hypothesis" coterie (see Lepore and Selleri 1990), most philosophers of physics agree that Bell's theorem and related results show that nature is nonlocal or "contextual" in some sense, assuming the predictions of quantum mechanics (QM) are more or less correct. But in what sense? Some, such as Home and Sengupta (1984) and Fine (1982), have claimed that Bell-type results have little to say about nonlocality. Furthermore, the dozens of clever new nonlocality proofs discovered over the past seven years (cf. Hardy 1993, Mermin 1990, Greenberger *et al.* 1990) rely on essentially the same deterministic locality assumptions used by Bell way back in 1964. To make useful philosophical headway, we must first establish, once and for all, that Bell-type results really do say something about nonlocality. Then, we must *weaken* the assumptions used to derive a nonlocality no-go theorem, thereby helping us to zero in on what locality assumptions must be renounced, and to cut off options previously available to "local realist" hidden-variable theorists such as Selleri. In this huge chapter, I will try to accomplish these goals.

First, in section 2.2, I help to establish that Bell-type derivations really do bear on the issue of nonlocality, by refuting Home and Sengupta's (1984) argument to the contrary. Then, in section 2.3, I criticize Heywood and Redhead's (1983) nonlocality proof on the grounds that it relies on too many assumptions. Section 2.4 reviews a cleaned-up version of Heywood and Redhead's proof. Then, in section 2.5, I present the first-ever "algebraic" nonlocality proof that relies on probabilistic (as opposed to

deterministic) locality assumptions. An "algebraic" nonlocality theorem invokes the perfect EPR-type correlations of QM, as opposed to the statistical correlations exploited by Bell-type inequalities. Section 2.6 cuts off another road previously available to hidden-variable theorists, by showing how theorem 2.5 can be modified so as to rule out theories that almost, but don't quite, reproduce the perfect correlations of QM. (Throughout this dissertation, I name theorems by the section in which they appear.) Finally, returning to the perfect correlations, I derive a nonlocality theorem using weakened locality assumptions. The result is, to my knowledge, the "best" nonlocality no-go theorem to date, in the sense of using the weakest locality assumptions.[1]

Sections 2.2 and 2.3 are nitpicky and dull, recommended only for those with a masochistic interest in nonlocality proofs. The real action starts in section 2.4.

[1]Bell-type results that attempt to use counterfactuals without assuming counterfactual definiteness, if valid, also use very weak assumptions. Comparing the "relative weakness" of my locality assumptions to (for instance) Stapp's (1993, 1994) locality assumptions is extremely difficult, because the assumptions take such different forms (conditional probabilities versus nested modal-logic counterfactuals). I won't attempt that project here.

Section 2.2: Nonlocality *is* the issue

§2.2.1. Introduction

Home and Sengupta argue that "contextuality," not nonlocality, is the philosophical lesson to be drawn from Bell-type inequalities. In this section, I refute their claim. But first, let me set the stage by summarizing another challenge to the "standard" interpretation of Bell inequalities.

Fine (1982) shows that if Bell-type inequalities are satisfied for a given set of events, then there exist well-defined joint probabilities for all pairs of events in question. For instance, if a Bell inequality is satisfied for events *a, b, c,* and *d,* then there exist well-defined joint probabilities p(*a,b*), p(*a,c*), and so on. But according to QM, some events are incompatible, and hence do not correspond to a well-defined joint probability. For instance, if operators **A** and **B** don't commute, then QM does not ascribe any meaning, much less a well-defined value, to the joint probability p(A=a, B=b), where "A=a" refers to observable A being measured to have definite value a. Using these facts, Fine argues that "what the Bell inequalities are all about" is making "well-defined precisely those probability distributions for noncommuting observables whose rejection is the very essence of quantum mechanics." For Fine, locality simply isn't the issue.

Svetlichny *et al.* (1988) refute Fine in several ways. First, they show that if probabilities are interpreted as relative frequencies (in the standard von Mises-Church sense), then a Bell inequality can be derived even if some joint probabilities are assumed *not* to exist. But more important, they point out that just because a set of assumptions (i.e., those leading to Bell inequalities) imply the mathematical well-definedness of certain joint probabilities, this does *not* mean that those joint probabilities correspond to physical reality. For instance, the Stapp-Eberhard-Redhead (see Redhead 1987) version

of a Bell inequality is derived without assuming joint probability distributions for quantum-mechanically incompatible events. Fine's proof shows that consistent, well-defined joint probabilities indeed exist for those incompatible events. But in the theories considered by Stapp and Eberhard, those joint probabilities are meaningless, mere mathematical artifacts without any physical content or significance. For this reason, the locality conditions used to derive Bell inequalities do not commit us to *physically-meaningful* joint probabilities for incompatible events. Bell inequalities are "about" locality, not about joint probability distributions.

Another challenge to the orthodox view that Bell's inequality is a *nonlocality* result comes from Home and Sengupta (henceforth H&S). They claim to derive a Bell-type inequality assuming only a noncontextual hidden-variable framework, thereby showing that no noncontextual theory reproduces the statistical predictions of quantum mechanics (QM). Specifically, H&S claim that their derivation, unlike usual Bell arguments, invokes no locality condition.

I show that their derivation assumes determinism. Also, H&S's "local noncontextuality" is a physically implausible restriction unneeded in Bell arguments. Furthermore, not only does their noncontextuality condition encode a weak locality assumption, but their inequality fails to rule out an important class of strongly-nonlocal theories. Upshot: The only "local" theories constrained by H&S's inequality satisfy implausible conditions not assumed in standard Bell arguments.

§2.2.2. Notation and Preliminary assumptions

Consider an electron in a $^2P_{1/2}$ state (orbital angular momentum equals 1, total angular momentum equals 1/2, in units of $\hbar$). The wave-function is

$$\Psi = (\sqrt{2}Y^k_{1,1}\otimes X^k_- - Y^k_{1,0}\otimes X^k_+)/\sqrt{3},$$

where $Y^k_{l,m}$ is the spatial spherical harmonic corresponding to total orbital angular momentum l and k component of orbital angular momentum m, and where spin states X^k_+ and X^k_- correspond to k component of spin up and down. Note that the spherical harmonics specify the electron's state *relative* to the nucleus; more on this point in section 2.2.5.

Let $\mathbf{l}_n$ and $\mathbf{S}_n$ denote the operators corresponding to the n component of orbital angular momentum and n component of spin, while l_n and S_n denote the corresponding observables. Let $l_n(t)$ and $S_n(t)$ be the values obtained upon measurement of l_n and S_n at time t, where we take $S_n = \pm 1$ instead of $\pm 1/2$.

Finally, let $L_n(t)$ be the "dichotomized" measured value of $l_n(t)$, where

(a) if $l_n(t)=0$, then $L_n(t)=+1$;

(b) if $l_n(t)=\pm 1$, then $L_n(t)=-1$.

Formally, $L_n=f(l_n)$, where $f(w)=1-2w^2$.

Notice that for all directions (unit vectors) **a** and **b**, $\mathbf{L}_a$ and $\mathbf{l}_a$ commute with $\mathbf{S}_b$, and hence l_a and S_b are commeasurable according to QM.

§2.2.3. Determinism

In this section, I trace H&S's derivation of a Bell-type inequality, revealing their determinism assumption.

Since $L_z(t)$, $L_a(t)$, and $S_b(t)$ all equal ± 1, H&S claim that inequality (1) follows:

$$(1) \qquad -L_z(t)L_a(t) + L_z(t)S_b(t) - L_a(t)S_b(t) \geq -1$$

The authors write: "Since we are considering the dispersion-free states to be noncontextual, it implies that if the outcome [of measuring] S_i is $S_i(t)$ in one pair, it will also be $S_i(t)$ in the other pair involving S_i." Inequality (1) necessarily holds, however, only under the following condition: If we measure S_b and l_z at time t and obtain $S_b(t)$, then $S_b(t)$ necessarily *would* have been obtained had we measured S_b and l_a instead. In other words, the result of measuring S_b must be counterfactually definite. As Redhead (1987, pp. 90-96) and others show, counterfactual definiteness relies on determinism. The following intuition underlies their formal arguments: Suppose measurement of l_z or l_a in no way affects measurement of S_b. Even so, we cannot assert what $S_b(t)$ *would* have been had conditions differed, unless we assume determinism. For, if the result of measuring S_b is truly indeterministic (irreducibly random), then measuring S_b and l_a might have yielded a different $S_b(t)$ than was obtained measuring S_b and l_z, *not because measurement of l_a or l_z disturbs measurement of S_b*, but simply because $S_b(t)$ is random and therefore might have come out differently. Therefore, in writing inequality (1), H&S assume determinism.

Here's another way to see that counterfactual definiteness enters in. The first term in inequality (1) has meaning only if measurement results of two incompatible (noncommuting) observables, L_z and L_a, are simultaneously well-defined. But if (say) L_z gets measured, then "L_a" can only mean the result we (counterfactually) *would* have obtained upon measuring L_a. Fortunately for H&S, since we can define those two values counterfactually under an assumption of determinism, it is irrelevant that we cannot measure them simultaneously.

Next, H&S note that, for state Ψ, if we let directions **k** and **z** coincide, then S_z and L_z are perfectly correlated: QM predicts that simultaneous measurement of S_z and l_z (or L_z) yields $S_z(t)=L_z(t)$ with probability one. H&S use this correlation to modify (1):

$$(2) \qquad -S_z(t)L_a(t) + L_z(t)S_b(t) - L_a(t)S_b(t) \geq -1$$

H&S write: "Summing over the relations of the [form (2)] over all the dispersion-free subensembles constituting the quantum mechanical ensemble and taking the average, we obtain the following inequality involving the quantum-mechanical expectation values:

$$(3) \qquad -\langle S_z L_a \rangle_\Psi + \langle L_z S_b \rangle_\Psi - \langle L_a S_b \rangle_\Psi \geq -1."$$

Quantum mechanics predicts a violation of (3) for suitably chosen angles between **a**, **b**, and **z**.

§2.2.4. Local noncontextuality

So far, I've shown that the H&S Bell-type derivation rests on determinism. I now argue that H&S's noncontextuality assumption is physically unmotivated.

H&S write that their derivation applies to all noncontextual theories. We must clarify the meaning of "noncontextual" in order to compare it to the locality assumption usually employed in Bell arguments. Generally, "noncontextuality" means the following:

Full noncontextuality: The result of a measurement on a system does not depend on which other simultaneous measurement(s) we perform on that same system or on a second system.

The locality condition used to derive a Bell inequality in a deterministic framework (cf. Redhead 1987) is

Bell locality: The result of a measurement on a system does not depend on which measurement(s) we perform on a second system, where measurement of the second system occurs spacelike separated from measurement of the first system.

Clearly, full noncontextuality implies Bell locality. Therefore, if H&S were to assume full noncontextuality, their derivation would lose much of its physical interest, because their assumptions (full noncontextuality and determinism) would be stronger than the usual deterministic Bell assumptions (Bell locality and determinism). H&S stress,

however, that their derivation refers to measurements associated with a single system "having no spatially separated components." That is, they claim to assume only

Local Noncontextuality : The result of a measurement on a system does not depend on which other simultaneous measurement(s) we perform on that same system.

Full Noncontextuality is equivalent to the conjunction of Local Noncontextuality and Bell Locality. To justify inequality (1), H&S invoke Local Noncontextuality: $S_b(t)$ may not depend on which orbital angular momentum component (l_z or l_a) undergoes measurement at t.

Although obeyed by QM itself, Local Noncontextuality is an unreasonable general restriction to place on hidden-variable theories. It is physically motivated only for theories in which the measurement result (or measurement-result probability) for a given observable depends entirely on the state of the system being measured. By contrast, consider Bohm's theory (cf. Bohm *et al.* 1987), in which measurement results for many observables depend both on the hidden-variable microstate of the measured system and on the hidden-variable state of the apparatus. In a "spinless" version of Bohm's theory, measuring the angular momentum of a system would inevitably involve disturbing its position in some manner, which is turn affects the outcome of "spin" measurements. (That's because "spin," in this version of the theory, is not an "internal" variable, but is instead the "byproduct" of the particle's position in relation to the magnets involved in measuring its spin. The details aren't worth dredging up here.) Since all these measurements take a finite amount of time to complete, the two measurement processes

would have to occur simultaneously or immediately after each other. In either case, the "disturbance" argument just given applies, even though the relevant quantum mechanical operators commute. By contrast, in "regular" QM, that disturbance argument could be sidestepped, at least in principle, for simultaneous or nearly-simultaneous measurement of commuting observables. In brief, Local Noncontextuality is physically unmotivated for theories in which the "microstate" of the apparatus matters.

Since we can derive a Bell inequality without assuming Local Noncontextuality (see Jarrett 1984), and since such derivations apply to theories incorporating apparatus-microstate dependence, the fact that H&S's inequality restricts only Local Noncontextuality-obeying theories lessens the physical interest of their considerations.

§2.2.5. Locality and the H&S derivation

In this section, I reveal H&S's implicit locality assumption. I also argue that any reasonable theory obeying H&S's noncontextuality (Local Noncontextuality) also obeys Bell locality.

Implicit locality. H&S, discussing inequality (3), write, "We have, therefore, an example--albeit in the form of a thought experiment--indicating the incompatibility of Bell's inequality with quantum mechanical predictions concerning simultaneous measurement of commuting observables associated with a system having no spatially separated components."

The "system" is the $^2P_{1/2}$ electron, while the "commuting observables" are a component of spin and a component of orbital angular momentum. I now argue that a

component of orbital angular momentum is a property not just of the electron, but of the whole system comprised of electron and nucleus.

Consider l_z, the z component of orbital angular momentum with respect to the nucleus. That observable is a function of the *relative* positions and *relative* momenta of the electron and the nucleus. Formally, $l_z = p_y \cdot x - p_x \cdot y$, where p_x is the *difference* between the electron's x component of momentum and the nucleus's x component of momentum. Similarly for p_y. Also, x is the difference between the electron's x coordinate and the nucleus's x coordinate. Similarly for y. In brief, a component of orbital angular momentum is not a property of the electron; it is a relational property of the electron and nucleus. Therefore, a measurement of the electron alone cannot, in practice or in principle, constitute a measurement of l_z (or L_z). To measure l_z, we must measure the system-as-a-whole, or some other object (such as an emitted photon) produced during an interaction of the electron with the nucleus.

For instance, we could measure l_z by taking spectra of a lithium ion in state Ψ. But the measured energy level "fixes" a component of the ion's center-of-mass velocity, because the Doppler shift of the spectral line establishes a component of the ion's velocity with respect to the lab frame. The point is, we cannot measure l_z without measuring some observable that "belongs" to the whole electron-nucleus system (in this case, the ion's center-of-mass velocity). I claim this conclusion does not depend on my choice of experimental procedure; *any* experiment to measure the relational property l_z inevitably involves a measurement (or disturbance or "fixing") of the ion-as-a-whole or of the nucleus, not just of the electron.

We now see the falsity of H&S's claim that a laboratory test of their inequality involves "measurement of commuting observables associated with a system having *no spatially separated components.*" H&S make this assertion to stress that their Bell argument, unlike its predecessors, does not assume locality. But in fact, H&S's noncontextuality assumption incorporates a weak locality condition. Their actual assumption is

H&S noncontextuality/locality: The result of measuring S_b does not depend on which measurement (l_z or l_a) we perform simultaneously on the electron-nucleus system-as-a-whole.

This is in part a weak locality condition, because part of the system-as-a-whole is the nucleus, which is spatially separated from the electron. The condition also incorporates Local Noncontextuality, because part of the system-as-a-whole is the electron itself. My argument that Local Noncontextuality is physically unreasonable therefore applies to H&S noncontextuality/locality.

Bell vs. H&S Locality. H&S's weak locality is less restrictive than Bell Locality. Nonetheless, H&S's inequality (3) restricts only the most implausible Bell-nonlocal theories, because only the most physically unreasonable theories obey H&S noncontextuality/locality but not Bell locality. Here's why:

H&S noncontextuality/locality demands that a measurement result on the electron not depend on the state of a measuring device that acts on a nearby system (i.e., the nucleus, or the electron-nucleus composite system). Bell locality demands that a

measurement result on the electron not depend on the state of a measuring device that acts on a *distant* (spacelike separated) system. Only a crafty hidden-variable theory will allow the states of distant devices, but not the states of nearby devices, to influence measurement results. For if some mechanism allows a distant apparatus to affect a measurement, then surely that same mechanism will allow a nearby apparatus to affect the measurement. In short, a theory obeying H&S noncontextuality/locality but not Bell Locality must assume the existence of a "field" that propagates instantaneously and acts at a distance, but never acts locally. Such a "field" seems utterly unbelievable. For instance, even the strong nuclear force binding quarks in a hadron does not *entirely* vanish at close range. Also, Bohm's "quantum potential" acts both locally and nonlocally.

In summary, not only do H&S implicitly assume a weak locality assumption, but any reasonable theory obeying H&S noncontextuality/locality also obeys Bell locality. Hence, H&S's inequality rules out only the most physically unreasonable Bell-nonlocal theories.

§2.2.6. Conclusion

First, I showed that H&S rely upon determinism. Then I discussed problems with their Local Noncontextuality, which standard Bell arguments do not use. H&S's argument rests not only on Local Noncontextuality but also on a weak locality assumption. Finally, we saw that only the most implausible Bell-nonlocal theories obey H&S's noncontextuality condition. Except for those unbelievable theories, therefore,

H&S's inequality only restricts deterministic Bell-local theories--the same theories restricted by standard Bell derivations in a deterministic framework.

Furthermore, as H&S admit, Kochen-Specker (1967) type results show the incompatibility of quantum mechanics with a broad class of noncontextual hidden-variable theories. Since H&S rely only on locally maximal observables (spin and orbital angular momentum), their derivation is slightly more general than Kochen and Specker's. But as argued above, their assumptions (which include Local Noncontextuality) are *less* general than those used in standard Bell arguments.

For these reasons, Home and Sengupta's derivation, though ingenious, does not set significant new limits on hidden-variable theorists, and does NOT displace locality as the central "issue" addressed by Bell inequalities.

Section 2.3: Extraneous assumptions: A case study

My argument of last section, combined with refutations to various other challenges (e.g., Fine vs. Svetlichny), establishes that Bell inequalities and kindred results have something to say about locality (or about noncontextuality, which is stronger than locality). Unfortunately, when formulating no-go theorems, we must be careful to invoke as few and as weak assumptions as possible. Otherwise, we can't narrow down what "flavor" of nonlocality nature violates.

In this section, I'll show how extraneous assumptions can scuttle an otherwise-impressive proof.

In 1983, Heywood and Redhead presented the first-ever "algebraic" nonlocality theorem. As noted above, "algebraic" theorems rely only on the perfect EPR-type correlations, not on the statistical correlations invoked by Bell-inequality derivations. (Such derivations were dubbed "algebraic" because most of them rely on descendants of Gleason's 1957 lemma and related results. Perhaps it's a misnomer, but let me retain it.) In section 2.4 below, I'll show why well-formulated algebraic proofs have certain philosophical advantages over "regular" Bell-type derivations. But here, I'll present a "case study" in the dangers of too many assumptions. My result is not particularly important, especially in light of sections 2.4 and 2.5, where I present a sleek algebraic nonlocality proof invoking the same EPR-type correlations as Heywood and Redhead used. Rather we should view this section as a warning buoy.

§2.3.1. Setting the stage: van Fraassen contextualism

In standard quantum mechanics, physical observables (other than mass, charge, and other quantities that are "fixed" for a given kind of particle) correspond 1:1 to Hermitian

operators. But alternative theories could violate this "Correspondence Principle." Bas van Fraassen (1973) suggests that multiple ontologically-distinct physical quantities may correspond to each *nonmaximal* (i.e., degenerate) Hermitian operator. (A maximal operator is one whose eigenstates span the relevant Hilbert space and whose eigenvalues are all distinct, i.e., no degeneracies.) As an example, suppose that maximal operators **A** and **B** don't commute, but nonmaximal **C** is a function of **A** and also a function of **B**: $\mathbf{C}=f(\mathbf{A})=g(\mathbf{B})$, where $[\mathbf{A},\mathbf{B}]\neq 0$. (In a spin-1 system, for instance, "**A**" could be the z component of spin S_Z, "**B**" could be the spin-Hamiltonian $\mathbf{H}_S=aS_x^2+bS_y^2+cS_z^2$, and "**C**" could be the square of the z component of spin, S_z^2.) We could measure physical observable in (at least) two different ways: by measuring A and applying function f to the result, or by measuring B and applying function g to the result. But since $[\mathbf{A},\mathbf{B}]\neq 0$, these two different measurement scenarios are mutually exclusive. Therefore, even in a deterministic framework, consistency does not require that the value of C found by measuring A would necessary equal the value of C found by measuring B. In symbols, $C_A\equiv f(A)$ need not equal $C_B\equiv f(B)$. Consistency with QM requires only that C_A and C_B measurement results display identical *statistical* distributions, not that C_A and C_B "agree" in individual cases. Using this formal fact, van Fraassen poses the possibility that C_A and C_B are ontologically distinct physical observables, every bit as "different" as position and momentum. Which of the many different "C's" is revealed by measurement depends on the *context* in which C is measured, i.e., on whether we measure C via A or C via B. For this reason, van Fraassen's construction is called "contextual." By contrast, a "noncontextual" theory or interpretation obeys the correspondence principle: observables correspond 1:1 to operators, and hence the "context of measurement" cannot affect the outcome.

§2.3.2. Introduction to Heywood-Redhead

Heywood and Redhead's algebraic nonlocality theorem relies on four explicit assumptions One of those principles, called FUNC*, is a contextualized version of FUNC, the algebraic constraint on observables' possessed values from which Kochen and Specker derive a contradiction. FUNC, which assumes a noncontextual setting, demands that the values of physical observables "mirror" the algebraic relations between the corresponding operators: if $\mathbf{Q}=f(\mathbf{R})$, then $Q=f(R)$. Although many kinds of hidden-variable theories violate FUNC, FUNC* seems so trivial that it defies analysis; indeed, in Heywood and Redhead's original prepublication draft, they didn't state FUNC* explicitly, but simply built it into their notation.[2] Because FUNC* is so obvious, however, we must understand its physical content. Otherwise, we risk smuggling in unanalyzed physical assumptions. Fine (1988) criticizes the Heywood-Redhead proof on these grounds: "The 'innocent looking' contextualized function condition [FUNC*] is not examined critically nor motivated physically. Indeed, it seems a purely formal constraint whose primary virtue is to make possible the demonstration of a no-go result."

In this paper, I explore the properties of hidden-variables theories obeying FUNC* and obeying the Value Rule, another of Heywood and Redhead's assumptions. Both principles follow in part from a version of Faithful Measurement, which requires measurement to "reveal" the value possessed by an observable. These considerations allow us to derive a Heywood-Redhead contradiction from physical (as opposed to purely formal) conditions, thereby clarifying which theories the contradiction rules out. We'll see that Heywood and Redhead make nontrivial "extra" assumptions not needed in

---

[2]Arthur Fine first pointed out the implicit reliance on FUNC*, and proved FUNC* to be both consistent with and independent of the Value Rule.

standard Bell derivations. Therefore, this particular "algebraic" nonlocality result does not improve upon Bell inequalities.

§2.3.3. Notation and preliminary assumptions

Consider a system comprised of two well-separated spin-1 particles. Following Heywood and Redhead, I restrict attention to operators and observables with discrete spectra. An observable is locally maximal iff its associated Hermitian operator is of the type $\mathbf{A}\otimes\mathbf{I}$ or $\mathbf{I}\otimes\mathbf{B}$ on the product Hilbert space $H_1\otimes H_2$, where H_1 (H_2) is the Hilbert space associated with particle 1 (2), where **I** is the identity operator, and where **A** (**B**) is maximal on H_1 (H_2).

To achieve as general a theorem as possible, Heywood and Redhead allow that the Correspondence Principle may fail. Therefore, as discussed above, many different ontologically-distinct physical quantities $\{Q_i\}$ corresponding to **Q** may exist. In Heywood and Redhead's "de-Ockhamized" framework, each member of $\{Q_i\}$ is associated with a different maximal observable. Let $Q_{(R)}$ be the member of $\{Q_i\}$ ontologically associated with maximal observable R. Let $[Q]_{(R)}(D,E)$ be the possessed value of $Q_{(R)}$ in an environmental context where measurement of D on particle 1 and measurement of E on particle 2 occur at time t. If X is maximal, then $[Q]_{(R)}(X)$ denotes the possessed value of $Q_{(R)}$, given that measurement of X occurs at t. Because Heywood and Redhead's notation is already so baggy, I won't indicate the time dependence of these possessed values.

If $\mathbf{A}\otimes\mathbf{I}$ and $\mathbf{I}\otimes\mathbf{B}$ are locally maximal, then $\langle\mathbf{A},\mathbf{B}\rangle$ denotes a maximal operator found by mathematically combining $\mathbf{A}\otimes\mathbf{I}$ and $\mathbf{I}\otimes\mathbf{B}$ in the appropriate manner. Strictly speaking, many different maximal operators can be forged from $\mathbf{A}\otimes\mathbf{I}$ and $\mathbf{I}\otimes\mathbf{B}$; but we'll assume that all these operators correspond to (functions of) a single physical

quantity. (Remember, even in this "contextual" framework, *maximal* operators and observables remain noncontextual, i.e., remain in 1:1 correspondence.)

In a hidden-variable theory, these possessed values depend on hidden parameters. Let λ denote the ontological (hidden-variable) state at time t of the two-particle system. Similarly, μ_Q is the hidden-variable microstate of an apparatus set to measure Q. More precisely, an apparatus in state μ_Q will "measure" one of the $\{Q_i\}$ corresponding to **Q**. Perhaps we don't even know which Q_i gets measured.

These λ and μ may evolve either deterministically or stochastically in time. State λ is "consistent with" quantum state ϕ iff a system described by quantum state ϕ can in principle occupy state λ. States λ, μ_Q, and μ_B are "consistent" iff a system in state λ can simultaneously interact with measuring devices in states μ_Q and μ_B. Hence, if μ_Q and μ_B are mutually consistent with some λ, then at least one of the $\{Q_i\}$ is commeasurable with one of the $\{B_i\}$.

Let $p(Q=q \mid \phi)$ be the probability, as calculated by QM, that a system in quantum state ϕ would yield value q upon measurement of Q. In QM, these probabilities are well-defined, since only one physical quantity Q corresponds to **Q**. In other words, QM is a noncontextual theory. Let $p(Q=q, R=r \mid \phi)$ be the QM joint probability that simultaneous measurement of Q and R would yield q and r, respectively.

Finally, let "Q=q" denote that measurement of Q at time t yielded result q. Of course, when we find that Q=q, we don't know which of the Q_i actually got measured.

§2.3.4. Heywood and Redhead's assumptions

As implied above, Heywood and Redhead assume all observables possess values, where many observables may correspond to a single nonmaximal operator. I assume this *realism of possessed value* throughout.

Next, Heywood and Redhead assume two locality principles:

Ontological locality (OLOC): $[Q]_{(<A,B>)}(D,E) = [Q]_{(<A,C>)}(D,E)$

[where Q is a locally maximal observable associated with particle 1].

Environmental locality (ELOC): $[Q]_{(R)}(D,C) = [Q]_{(R)}(D,E)$

[where Q is associated with particle 1].

ELOC expresses the idea, motivated in part by relativity theory, that a property of a particle (e.g., a possessed value associated with that particle) cannot depend on the setting of an apparatus well separated from that particle. Redhead (1987) and others show that ELOC implicitly rests on counterfactual definiteness, and therefore applies unproblematically only to deterministic theories.

OLOC requires that observables not be "split" by the ontological context associated with a separated system. The *local* ontological context may split observables; in general $[Q]_{(<A,B>)}(D,E) \neq [Q]_{(<C,B>)}(D,E)$, because $Q_{(<A,B>)}$ and $Q_{(<C,B>)}$ are different physical quantities. OLOC requires that such splitting occur only with respect to the local context: $Q_{(<A,B>)}$ and $Q_{(<A,C>)}$ denote the *same* observable when Q is associated with particle 1. See Redhead (1987) for more discussion. Taken together, OLOC and ELOC encode the same locality assumptions used in standard Bell derivations in a deterministic framework.

But Heywood and Redhead also assume two auxiliary conditions:

Value Rule: For maximal R, $p(R=r \mid \phi)=0 \rightarrow [R]_{(R)}(R) \neq r$.

FUNC:* if **R** is maximal and **Q**=f(**R**), **D**=g(**R**), and **Q**=h(**D**), then

$[Q]_{(R)}(R) = h([D]_{(R)}(R))$.

Value Rule requires maximal observables not to possess values "ruled out" by the QM formalism. FUNC* requires that, with respect to a given ontological context, the values of observables mirror the functional relationships between the corresponding operators. In other words, FUNC* demands that *within a given ontological context*, FUNC must hold. The physical significance of these two axioms, especially FUNC*, requires explication beyond that provided by Heywood and Redhead.

§2.3.5. Physical Significance of FUNC and Value Rule*

In this section, I derive FUNC* and VR from three physical conditions, one of which is a contextualized version of Faithful Measurement. This helps me to explicate the physical assumptions "hiding" in FUNC*.

First, I present and briefly discuss my three assumptions. According to QM, observables associated with commuting operators are commeasurable. I impose a restricted version of this requirement:

Commeasurability : If **Q**=f(**R**) for maximal **R**, then for any λ there exist consistent apparatus microstates μ_Q, μ_A, μ_B, etc., such that a system in state λ, upon interacting with apparatuses in states $\{\mu_Q, \mu_A, \mu_B...\}$, would yield measured values for Q and R.

Commeasurability asserts that it is possible to measure Q in conjunction with a compatible maximal observable R. More precisely, according to Commeasurability, at least one of the $\{Q_i\}$ is commeasurable with R. Commeasurability does not specify

which of the $\{Q_i\}$ is measured by this arrangement. Commeasurability also requires that such joint measurements yield joint results. This condition does not demand that all pairs of observables associated with commuting operators be commeasurable. Nonetheless, Commeasurability could fail for some prism model theories, in which essential detector inefficiencies prevent certain joint measurements from always yielding results; see Fine (1989).

My second assumption is Faithful Measurement. In the contextual theory under consideration, measurement of Q could conceivably reveal $[Q]_{(R)}(R)$, $[Q]_{(X)}(R)$, or the value of some other $\{Q_i\}$. My version of Faithful Measurement requires that the member of $\{Q_i\}$ picked out by measurement be such that the ontological "context" matches the environmental context:

Faithful Measurement: If $\mathbf{Q}=f(\mathbf{R})$ for maximal **R**, then simultaneous measurement of Q and R at time t, or measuring R alone and then applying function f to the result, would necessarily yield the value $[Q]_{(R)}(R)$ for Q.

According to Faithful Measurement, measurement of Q in conjunction with measurement of (maximal) R reveals the value of $Q_{(R)}$. Notice that Faithful Measurement rests on Commeasurability by assuming the commeasurability of Q and R. Only when we measure Q together with a maximal observable does Faithful Measurement constrain which member of $\{Q_i\}$ is revealed.

Faithful Measurement fails for theories such as David Bohm's, in which measurement results for some observables depend on the hidden-variable "microstate" of the measuring apparatus. In such theories, measurement does not simply reveal a

property of the particle, because microproperties of the apparatus affect the measurement outcome.

Since Faithful Measurement requires that measurement *necessarily* reveal some property of the particle, this condition can apply only to deterministic theories. But as we saw earlier, the Heywood-Redhead proof applies unproblematically only to deterministic theories, due to ELOC; so Faithful Measurement simply exploits a determinism assumption already implicit in the theorem.

My final assumption is

Measured Value Rule: If $\mathbf{Q}=f(\mathbf{R})$ for maximal **R**, then

$p(R=r, Q=q \mid \phi)=0 \rightarrow R\neq r$ or $Q\neq q$.

Measured Value Rule, which also rests on Commeasurability, requires the nonoccurrence of certain joint measurement results "ruled out" by QM. Which hidden-variable theories violate Measured Value Rule? Suppose states λ, μ_R, and μ_Q are consistent and ϕ-consistent. Then the set of states $\{\lambda, \mu_R, \mu_Q\}$ is "anomalous" if a system in state λ, upon interacting with apparatuses in states μ_Q and μ_R, would yield $Q=q$ and $R=r$ even though $p(R=r,Q=q \mid \phi)=0$. Thus, a deterministic theory violates Measured Value Rule just in case it contains anomalous hidden-variable states. But such a theory does not reproduce the statistical predictions of QM unless the anomalous states are a zero-measure subset of all accessible particle and apparatus states. In brief, two types of theories violate Measured Value Rule: (a) those violating QM, such as those proposing small "corrections" to QM; and (b) those incorporating the seemingly *ad hoc* feature of anomalous hidden-variables states in zero-measure sets.

Derivation of FUNC and Value Rule.* I now derive FUNC* and Value Rule from the three conditions just presented.

According to Faithful Measurement, measurement of maximal R reveals $[R]_{(R)}(R)$; to prove this, let **Q=R** in the definition of Faithful Measurement. Before proceeding, I must discuss how to measure maximal R. Intuitively, one method is to measure observables associated with particles 1 and 2 separately. For instance, if R=<A,B>, then we expect that a way to measure R is simultaneously to measure A and B. I prove in the appendix to this section that if **Q** is an operator associated with particle 1 and **Q**=f(**R**) for maximal **R**, then there exist an infinite number of pairs of commuting operators {(**Q'**,**B**)} such that R=<A,B>, where **A**=**Q**+**Q'** and where **Q** and **Q'** commute. I assume that for at least one of these (Q',B) pairs, a way to measure R is to measure Q, Q', and B simultaneously. I do *not* assume that for any arbitrary (Q',B) pair in {(Q',B)}, a way to measure R is to measure Q, Q', and B. Nor do I make assumptions about how to calculate the measurement result R from the measurement results Q, Q', and B. For instance, I do not require that the measurement results obey A=Q+Q'. I assume only that for given Q and R, there exist Q' and B such that a simultaneous measurement of Q, Q', and B constitutes a measurement of R. More precisely, I assume that for some **Q'** and **B**, there exists a member of $\{Q'_i\}$, a member of $\{B_i\}$, and a member of $\{Q_i\}$ such that simultaneous measurement of those three quantities constitutes a measurement of R. Please regard this assumption as an extension, or perhaps a clarification, of Commeasurability. As we have seen, a large class of prism models and Bohm-type theories may violate this condition. Such theories escape the following proof:

Theorem: Faithful Measurement & Commeasurability & Measured Value Rule → FUNC* & Value Rule.

Proof. First I derive FUNC*. Suppose $\mathbf{Q}=f(\mathbf{R})$ for maximal **R**. Consider an experimental arrangement that simultaneously measures Q, Q', and B, where that tri-joint measurement constitutes a measurement of R. (By the extension of Commeasurability just discussed, such a Q' and B exist.) For all real numbers r_i and all quantum states ϕ, $p(Q\neq f(r_i), R=r_i \mid \phi)=0$. By Measured Value Rule, it follows that for all r_i, $R\neq r_i$ or $Q=f(r_i)$. But $R=r$ for some real number r, by Commeasurability. Therefore, $Q=f(r)=f(R)$. In words, the Q measurement result is the "correct" function of the R measurement result.[3]

As noted above, the experiment under consideration constitutes a measurement of R. Hence, Faithful Measurement demands that $R=[R]_{(R)}(R)$. The arrangement also incorporates a measurement of Q. Since Q and R undergo simultaneous measurement, Faithful Measurement demands that the measurement result Q equal the appropriate possessed value: $Q=[Q]_{(R)}(R)$. Because $Q=f(R)$, it follows that $[Q]_{(R)}(R)=f([R]_{(R)}(R))$.

By equivalent reasoning, if $\mathbf{D}=g(\mathbf{R})$, then $[D]_{(R)}(R)=g([R]_{(R)}(R))$. So, if $\mathbf{Q}=h(\mathbf{D})=h\circ g(\mathbf{R})$, then $[Q]_{(R)}(R)=h\circ g([R]_{(R)}(R))=h([D]_{(R)}(R))$. This is just FUNC*.

Now I derive Value Rule. As before, let $\mathbf{Q}=f(\mathbf{R})$ for maximal **R**. Suppose $p(R=r \mid \phi)=0$. Then for all real numbers q_i, $p(Q=q_i, R=r \mid \phi)=0$. Measured Value Rule therefore requires that for all real numbers q_i, $R\neq r$ or $Q\neq q_i$. But $Q=q$ for some real number q, by Commeasurability. Therefore $R\neq r$. Since Faithful Measurement demands that $R=[R]_{(R)}(R)$, it follows that $[R]_{(R)}(R)\neq r$. *Q.E.D.*

§2.3.6. Conclusion

---

[3]This argument resembles a proof given by Fine (1974).

The fact that FUNC*, Value Rule, OLOC, and ELOC entail a Kochen-Specker algebraic contradiction, along with theorem 2.3.5, implies

Theorem : Faithful Measurement & Commeasurability & Measured Value Rule & OLOC & ELOC → Kochen-Specker contradiction,

where ELOC, as well as Faithful Measurement, rests on determinism. This theorem clarifies the physical interpretation of Heywood and Redhead's result by pinning down which classes of theories the auxiliary assumptions rule out. Specifically, at least three classes of deterministic local (OLOC and ELOC obeying) theories incorporating possessed values escape the Heywood-Redhead argument:

(a) Those violating Faithful Measurement. Some such theories incorporate the Bohm-like feature of allowing measurement results to depend on measuring-device microstates.

(b) Those violating Measured Value Rule. Such theories either violate QM's statistical predictions or incorporate anomalous states in zero-measure sets.

(c) Those violating Commeasurability. Such theories rule out the possibility of certain joint measurements permitted by QM, or at least allow those joint measurements not to yield joint results. Prism models disobey Commeasurability. Neither Faithful Measurement nor Measured Value Rule makes sense if Commeasurability fails.

Many hidden-variables theories violate Faithful Measurement. Logically, of course, a theory could disobey Faithful Measurement while obeying FUNC* and Value Rule, in which case the Heywood-Redhead result still applies. But if Faithful

Measurement fails, then FUNC* loses its physical motivation, for this reason: FUNC* requires the functional relationships between possessed values to mirror the functional relationships between the underlying operators, which in turn (according to QM) establish functional relationships between joint measurement results. But if possessed values do not correspond to measurement results, then we have little reason to suppose that the functional relationships between possessed values should mirror the functional relationships between measurement results. Hence, if Faithful Measurement fails, then FUNC* becomes (as Fine writes) an *ad hoc* formal constraint instead of a physically motivated principle.

In conclusion, it would be an improvement to derive a Heywood-Redhead type contradiction without assuming FUNC* or any such extraneous assumptions. The rest of this chapter will accomplish exactly that.

§2.3.7. Appendix to Section 2.3

I show that if $\mathbf{Q}=f(\mathbf{R})$ for maximal $\mathbf{R}$, and if $\mathbf{Q}$ is associated with particle 1, then there exist an infinite number of Hermitian operators $\mathbf{Q'}$ associated with particle 1 and $\mathbf{B}$ associated with particle 2 such that $R=\langle A,B\rangle$, where $\mathbf{A}=\mathbf{Q}+\mathbf{Q'}$.

As throughout, I consider observables with discrete spectra. Recall that $\mathbf{R}$ is maximal on the two-particle system iff $\mathbf{R}=\sum_{ij} r_{ij}\mathbf{P}_i\otimes\mathbf{P'}_j$, where the projection operators $\mathbf{P}_i$ ($\mathbf{P'}_j$) form a complete orthonormal basis for the operator Hilbert space associated with particle 1 (2), and where the $\{r_{ij}\}$ are all "distinct." (A set of numbers is "distinct" iff all of them are different.) Since $\mathbf{Q}=f(\mathbf{R})$ and since $\mathbf{Q}$ is associated with particle 1, $\mathbf{Q}=\sum_i q_i\mathbf{P}_i$, where the $\{q_i\}$ are not necessarily distinct. Now consider two operators $\mathbf{A}=\sum_i a_i\mathbf{P}_i$ and $\mathbf{B}=\sum_j b_j\mathbf{P'}_j$, where the $\{a_i\}$ and $\{b_j\}$ are distinct. Then by definition $\mathbf{A}$ and $\mathbf{B}$ are locally maximal on particles 1 and 2, respectively. As shown in Heywood

and Redhead, $R=\langle A,B\rangle$ iff $\mathbf{R}=\sum_{ij}F(a_i,b_j)\mathbf{P}_i\otimes\mathbf{P'}_j$, where function F is 1:1 over the relevant domain, the ordered pairs (a_i,b_j). But such a function exists, namely $F(a_i,b_j)=r_{ij}$. To see that this function is 1:1, notice that since the $\{a_i\}$, $\{b_j\}$, and $\{r_{ij}\}$ are distinct; $F(a_i,b_j)=F(a_g,b_h)$ iff $i=g$ and $j=h$. So F is 1:1, and therefore $R=\langle A,B\rangle$. Now just let $\mathbf{Q'}=\mathbf{A}-\mathbf{Q}$; that is, $\mathbf{Q'}=\sum_i(a_i-q_i)\mathbf{P}_i$. Note that **Q** and **Q'** commute, because $\mathbf{Q'}=k(\mathbf{Q})$, where $k(q_i)=a_i-q_i$. Hence, I have shown that there exist Hermitian operators **Q'** associated with particle 1 and **B** associated with particle 2 such that $R=\langle A,B\rangle$, where $\mathbf{A}=\mathbf{Q}+\mathbf{Q'}$. In fact, I have demonstrated the existence of an infinity of such **Q'** and **B**, because there are an infinite number of $\mathbf{A}=\sum_i a_i\mathbf{P}_i$ and $\mathbf{B}=\sum_j b_j\mathbf{P'}_j$ such that the $\{a_i\}$ and $\{b_j\}$ are distinct.

Section 2.4: Gleason's lemma and nonlocality

Here, I set the stage for my new proofs in section 2.5. First, I'll briefly review how a Gleason-Kochen-Specker algebraic contradiction arises. Then I'll present a version of Brown and Svetlichny's (1990) algebraic nonlocality theorem, which builds upon Stairs (1983). Like Heywood and Redhead, Brown and Svetlichny invoke the Kochen-Specker contradiction. But unlike Heywood and Redhead, Brown and Svetlichny assume nothing more than the standard deterministic locality conditions used in Bell derivations. That is, Brown and Svetlichny find a way to dispense with the kinds of extraneous assumptions I criticized in the previous section. I present his nonlocality proof because it's interesting (and remarkably simple) in its own right, because I played a small role in helping to develop it, and because it naturally leads into my own theorems.

§2.4.1. Gleason and descendants

Consider the unit sphere. Unit vectors correspond 1:1 to points on that sphere. So, I can uniquely specify a point by specifying a unit vector. An *orthogonal triad* is a set of three points corresponding to three mutually orthogonal unit vectors. So for instance, {north pole, equator at 20° longitude, equator at 110° longitude} is an "orthogonal triad" on the Earth's surface. Notice that any given point is a member of many (indeed, an infinite number of) orthogonal triads.

Can we paint the Earth's surface red and blue such that, within any orthogonal triad, two points are red and one point is blue?[4] Surprisingly, as Gleason (1957) first proved, the answer turns out to be "no."

---

[4]I owe this formulation of the problem to Michael Redhead (1987).

Bell (1966), and later Kochen and Specker (1967), showed that a Gleason type result ensues even when only a finite number of points are considered. Specifically, Kochen and Specker consider 43 orthogonal triads, and show that they can't be "painted" as specified above. Those 43 orthogonal triads consist of only 117 points instead of 43x3=129, because several points "belong" to more than one triad. Peres (1990) improves on the Kochen-Specker result by showing that only 16 triads (consisting of 33 points) need be considered to reach the same conclusion.

Bell, and independently Kochen and Specker, realized that Gleason-type results can be used to rule out certain hidden-variable theories. For instance, we can quickly dissolve a whole class of noncontextual theories that assign definite values to all observables associated with Hermitian operators. I'll do so now.

Kochen-Specker theorem. Let [Q] denote the possessed value of observable Q. Consider "noncontextual" theories, according to which the (possessed or measured) value of a nonmaximal observable[5] does not depend on the "context" in which the observable is measured. More formally, in the noncontextual theories considered here, if nonmaximal operator **A**=f(**B**)=g(**C**), where **B** and **C** are commuting *or noncommuting* maximal observables, there exists only one physical quantity A corresponding to **A**. So, there's no van Fraassen style ontological splitting. And furthermore, we can calculate the value of A by taking the relevant function of [B] *or* [C]. Formally, [A]=f([B])=g([C]). In brief, I've assumed a condition Kochen and Specker call

FUNC: For all **A** and **B**, If **A**=f(**B**), then [A]=f([B]).

---

[5]When I call an observable "maximal" or "nonmaximal," I'm really referring to the operator associated with that observable.

This condition seems innocuous. It simply says, for instance, that a particle's kinetic energy can be calculated by squaring the value of its speed and multiplying by m/2: $[K]=m[v]^2/2$. For noncontextual theories that assign values to all observables corresponding to Hermitian operators, FUNC must hold to insure that the possessed values obey the "right" algebraic relationships.

As Kochen and Specker show, however, FUNC implies a Kochen-Specker ("coloring") contradiction. I'll present a quick version of their argument due to Redhead. Consider a spin-1 particle. So, the quantum number s equals 1. From QM, the particle occupies an eigenstate of the S^2, with eigenvalue s(s+1)=2. So, $[S^2]=2$. But of course, $S^2=S_x^2+S_y^2+S_z^2$. FUNC therefore implies

$$2 = [S^2] = [S_x^2]+[S_y^2]+[S_z^2] = [S_x]^2+[S_y]^2+[S_z]^2$$

For a spin-1 particle, $[S_n]=-1, 0,$ or 1.[6] Therefore, $[S_n]^2=0$ or 1. Consequently, of the three values $\{[S_x]^2, [S_y]^2, [S_z]^2\}$, two of them must equal 1 while the third must equal 0. Otherwise, those three values couldn't add up to 2. Furthermore, since $S^2=S_x^2+S_y^2+S_z^2$ for *any* orthogonal triad $\{x, y, z\}$, this conclusion applies to the spin-components along *any* orthogonal triad of directions. Formally, for any orthogonal triad $\{x, y, z\}$, the values $\{[S_x]^2, [S_y]^2, [S_z]^2\}$ must be such that two of them equal 1 while the third equals 0.

---

[6]I've just implicitly assumed the "Spectrum Rule," according to which the possible values of an observable are simply the spectrum of the corresponding operator. Since QM implies that measurement of an observable yields a value in the spectrum of the corresponding operator, violation of the Spectrum Rule would allow observables to possess values that don't correspond to measurement results. In other words, the measured value wouldn't necessarily equal the pre-existing possessed value. But what's the point of introducing "possessed values" (for all observables) if they aren't the values "revealed" by measurement?

To hook this up to Gleason's lemma and its descendants, color the unit sphere according to following scheme:

if $[S_n]^2=1$, then paint the point corresponding to **n** red.

if $[S_n]^2=0$, then paint the point corresponding to **n** blue.

But as discussed above, this is impossible. We've reached a contradiction. So, we must give up at least one of the assumptions. Specifically, we must abandon noncontextual deterministic theories that assign definite values to all observables consistent with FUNC.

Notice that locality didn't enter into this no-go theorem. On the other hand, it only rules out theories satisfying a particularly strong (and according to Bell, implausible) set of conditions. I won't get into this debate here.

A quick technical point: Kochen and Specker's original proof doesn't use S^2. Instead, it uses the "spin-Hamiltonian", $\mathbf{H}_S=aS_x^2+bS_y^2+cS_z^2$, where a, b, and c are all different. The details aren't worth reproducing. As Kochen and Specker discuss, if a spin-1 system (such as a helium atom in the "right" state) is placed in a weak magnetic field of rhombic symmetry, then $\mathbf{H}_S$ corresponds to its energy (or more precisely, the "part" of its energy due to the interaction between spin and magnetic field). So, even if you don't think all Hermitian operators correspond to "real" physical quantities, you have to admit that H_S does. In fact, Kochen and Specker lay out a complex scheme for measuring H_S. The fact that this measurement is extremely difficult to carry out in practice does not threaten the "validity" of H_S as a real physical observable.

§2.4.2. Gleason's descendants meet nonlocality

Heywood and Redhead, and independently Stairs, were the first to realize that a Gleason-Bell-Kochen-Specker contradiction could be used in a nonlocality proof. As discussed above, Heywood and Redhead rely on extraneous assumptions that cloud the physical interpretation of their theorem. And Stairs' result is more a suggestion for a possible proof than a fully formalized theorem. But more recently, Brown and Svetlichny (1990) formalized Stairs' outline into a rigorous proof. Notably, Brown and Svetlichny assume the same deterministic locality conditions invoked in standard Bell derivations.

Assumptions. I'll now lay out those conditions. We'll consider an EPR-type experiment in which two particles created at a common source speed in opposite directions and get measured at spacelike separation. I'll call the two particles "1" and "2."

First, assume that measurement results are fully determined by the state of the particles. So, this theorem does not address theories in which the apparatus "microstate" plays a role. Let [Q⊗I] denote the value that would be obtained if Q were measured on particle 1. [I⊗Q] denotes the analogous value of particle 2. Physically, [Q⊗I] is *determined* by the fully specified (hidden-variable) state of the particles. Although my notation leaves out the time dependence of [Q⊗I], such dependence is certainly allowed.

Second assume "Bell locality": [Q⊗I] may not depend on which observable gets measured on particle 2, and [I⊗Q] may not depend on which observable gets measured on particle 1. Expressed counterfactually, Bell locality demands that if we measure Q on particle 1 and R on particle 2, then we'll get the same [Q⊗I] as *would* have been obtained had we measured R' on particle 2. The implicit counterfactual definiteness here is not problematic because we've already assumed determinism. Clearly, this condition rules out a direct faster-than-light causal connection between the two measurements. I've

implicitly assumed that if an experimenter can control a measurement result on particle 1 by changing a setting on apparatus 2, then the nonlocal connection between those two "wings" of the experiment is causal. For a more careful discussion of causality, see chapter 5.

Finally, assume "Particle Locality": If the measuring apparatus "settings" are chosen at time t, then the state of the particles at time t does not depend on that choice of settings. Again, this condition rules out an instantaneous influence between two spacelike separated events, in this case the manipulation of an apparatus and the state-evolution of a particle that hasn't yet reached that apparatus.

Finally, I'll assume the Spectrum Rule, according to which a measurement result on Q must equal one of the eigenvalues of **Q**. As noted above, if this condition fails (for a nonzero-measure set of hidden-variable states), then the hidden-variable theory violates QM even before consideration of locality are brought in.

Theorem: Determinism & Bell Locality & Particle Locality & Spectrum Rule & (no apparatus hidden variables) → Contradiction with QM's perfect correlations.

Proof: I will now show that those assumptions contradict the perfect anticorrelations of QM.

Consider two spin-1 particles in their singlet state,

$$|\Psi_{singlet}\rangle = -3^{-1/2}(|S_x=0\rangle\otimes|S_x=0\rangle - |S_y=0\rangle\otimes|S_y=0\rangle + |S_z=0\rangle\otimes|S_z=0\rangle)$$

On particle 1, we'll measure the spin-Hamiltonian $H_S=aS_x^2+bS_y^2+cS_z^2$, while on particle 2 we'll measure one of corresponding spin components, either S_x, S_y, or S_z.

The eigenvalues of $\mathbf{H}_S$ are $\{h_X=b+c, h_Y=a+c, h_Z=a+b\}$. Quantum mechanics predicts the following perfect anticorrelations:

(a) $p(H_S\otimes I=h_X, I\otimes S_X=\pm 1 \mid \Psi_{singlet})=0$,

(b) $p(H_S\otimes I=h_X, I\otimes S_Y=0 \mid \Psi_{singlet})=0$,

(c) $p(H_S\otimes I=h_X, I\otimes S_Z=0 \mid \Psi_{singlet})=0$.

In this notation, $p(Q\otimes I=q, I\otimes R=r \mid \Psi)$ is the probability according to QM that, when the particles occupy quantum state Ψ, simultaneous measurement of Q on particle 1 and R on particle 2 would yield q and r, respectively.

Particle locality ensures that no matter which observables get measured, the same distribution of hidden-variable states underlie the quantum state $\Psi_{singlet}$. If the hidden-variable theory is to reproduce a given perfect correlation, then a measure-1 set of the hidden-variable states underlying quantum state $\Psi_{singlet}$ must mirror that perfect correlation. This proof will consider a finite number of perfect correlations--specifically, 12 correlations for each of the 16 orthogonal triads used in the Peres-Kochen-Specker proof, for a total of 192 perfect correlations. Since the intersection of a finite collection of measure-1 sets is itself a measure-1 set, there exists a measure-1 set of hidden-variable states that mirror *all* the perfect correlations considered here.[7] From now on, I'll consider the values associated with a hidden-variable state in that "intersection" set.

Suppose that $[H_S\otimes I]=h_X$. Since the hidden-variable state under consideration reproduces all the perfect correlations considered in this proof, we have (from above)

---

[7]That last bit of reasoning would not have been possible if my proof considered an infinite number of perfect anticorrelations. I'm taking advantage of the fact that the Kochen-Specker algebraic contradiction, unlike Gleason's original lemma, relies on a finite number of orthogonal triads. To my knowledge, this is the first proof to exploit that fact.

(a) $[H_S \otimes I] \neq h_X$ or $[I \otimes S_X] \neq \pm 1$,

(b) $[H_S \otimes I] \neq h_X$ or $[I \otimes S_Y] \neq 0$,

(c) $[H_S \otimes I] \neq h_X$ or $[I \otimes S_Z] \neq 0$.

By supposition, $[H_S \otimes I] = h_X$. So,

(a) $[I \otimes S_X] \neq \pm 1$,

(b) $[I \otimes S_Y] \neq 0$,

(c) $[I \otimes S_Z] \neq 0$.

Since $[I \otimes S_X] \neq \pm 1$, the Spectrum Rule implies $[I \otimes S_X] = 0$. And since, $[I \otimes S_{Y,Z}] \neq 0$, the Spectrum Rule implies $[I \otimes S_{Y,Z}] = \pm 1$. In summary, we have $[I \otimes S_X] = 0$, $[I \otimes S_Y] = \pm 1$, and $[I \otimes S_Z] = \pm 1$.

Now of course, that conclusion rests on the provisional assumption that $[H_S \otimes I] = h_X$. If we had supposed instead that $[H_S \otimes I] = h_Y$, equivalent reasoning (using analogous quantum perfect anticorrelations) would have given us $[I \otimes S_X] = \pm 1$, $[I \otimes S_Y] = 0$, and $[I \otimes S_Z] = \pm 1$. And had we supposed $[H_S \otimes I] = h_Z$, the same reasoning would have yielded $[I \otimes S_X] = \pm 1$, $[I \otimes S_Y] = \pm 1$, and $[I \otimes S_Z] = \pm 0$. In summary, no matter what the spin-Hamiltonian equals, the three values $\{[I \otimes S_X]^2, [I \otimes S_Y]^2, [I \otimes S_Z]^2\}$ are such that two of those values equal 1, while the third equals 0.

Since $\Psi_{singlet}$ is spherically symmetric, the same reasoning--and hence the same conclusion--applies to any orthogonal triad of directions. (Remember, the hidden-variable state under consideration reproduces all 192 relevant perfect anticorrelations.)

Formally, for any of those 192 orthogonal triads, {**x'**, **y'**, **z'**}, the values {$[I\otimes S_{x'}]^2$, $[I\otimes S_{y'}]^2$, $[I\otimes S_{z'}]^2$} are such that two of those values equal 1, while the third equals 0.

Here's the punch line: For every unit vector **n**, try to "color" the corresponding point on the unit sphere with the value $[I\otimes S_n]^2$. As just shown, the result is such that for each orthogonal triad, two points are "1" (red) and the third point is "0" (blue). This is impossible, by the Peres-Kochen-Specker theorem. We've reached a contradiction. Therefore, any theory consistent with the above assumptions cannot reproduce the perfect correlations of quantum theory. *Q.E.D.*

You may wonder where Bell Locality entered into the reasoning. Well, some of the unit vectors involved in the Peres-Kochen-Specker theorem "belong" to more than one orthogonal triad. Suppose **x** is one of them. Then the above theorem ends up invoking not just $p(H_s\otimes I=h_x, I\otimes S_x=\pm 1 \mid \Psi_{singlet})=0$, but also $p(H_s'\otimes I=h_x, I\otimes S_x=\pm 1 \mid \Psi_{singlet})=0$, where $\mathbf{H_S'}=aS_x^2+bS_{y'}^2+cS_{z'}^2$, with {**x**, **y'**, **z'**} another orthogonal triad involving **x**. The value assigned to $I\otimes S_x$ by virtue of its correlation with $H_s\otimes I$ must equal the value assigned to $I\otimes S_x$ by virtue of its correlation with $H_s'\otimes I$, or else $[I\otimes S_x]$ isn't uniquely defined and the proof falls through. So, the value $[I\otimes S_x]$ associated with particle 2 may not depend on whether H_s or H_s' is measured on particle 1. This is guaranteed by Bell Locality.

§2.4.3. Summary

In this section, I introduced Gleason's lemma and its descendants, which establish the impossibility of mapping values to unit vectors (i.e., to points on the unit sphere) consistent with the "coloring rule" discussed above. I then showed how this mathematical result can be used to rule out a class of hidden-variable theories.

Unfortunately, Kochen and Specker's hidden-variable no-go theorem rules out only those theories satisfying a particularly strong set of assumptions. Finally, I showed how a Gleason-type contradiction can be employed in an algebraic nonlocality theorem that invokes the same deterministic locality conditions used in Bell derivations.

In the next section, I'll improve upon Brown and Svetlichny's result by deriving an algebraic nonlocality theorem in an indeterministic framework.

Section 2.5: Generalization of algebraic nonlocality proof to indeterministic setting

The most general Bell derivations employ locality assumptions that are weaker than those of Brown and Svetlichny in two major ways. First, as Clauser and Horne (1974) showed, by invoking a probabilistic condition called "Factorizability" we can avoid the assumption of determinism. Second, Bell-type derivations (cf. Jarrett 1984) can allow measuring apparatus "microstates" to affect measurement outcomes, as happens in Bohm's theory.

In this section, I'll show that the *weakest* probabilistic assumptions needed to derive a Bell inequality can be used to derive an algebraic nonlocality proof. By "algebraic," I mean a proof that invokes only the perfect (anti)correlations, as opposed to the more general statistical correlations, of QM. Then I'll spell out some of the philosophical advantages of this approach. (Later on, in section 2.7, I'll rederive my result from weakened locality conditions, conditions from which a Bell inequality cannot be derived.) But first, I must review in some detail the precise conditions needed in Bell-type derivations.

§2.5.1. Notation and preliminaries

In a standard EPR-type arrangement, Let A, A', etc., denote physical quantities that apparatus 1 can measure, while B and B' denote quantities that apparatus 2 can measure. In other words, A and A' are possible settings of apparatus 1. Notice that I'm streamlining the notation by writing A instead of A⊗I and B instead of I⊗B.

Let λ denote the ontological (fully specified) state of the pair of particles immediately before the particles undergo measurement. In my terminology, the

measurement begins when apparatus 1 (2) first starts to interact locally with particle 1 (2). Importantly, λ does *not* denote the state of the particles at the source, for reasons presented below.

Let μ_A denote the ontological state of an apparatus set to measure A. Call μ_A the "apparatus microstate." In general, many different microstates are accessible to an apparatus macroscopically set to measure A. According to some "micro-contextual" theories, measurement outcomes depend not just on the apparatus settings, but also on these microstates. The λ and μ states may evolve either deterministically or stochastically in time.

Throughout this dissertation, I use standard conditional probability notation: $p(b|a)$ is the objective probability of *b* given *a*, and $\rho(b|a)$ is the probability density of *b* given *a*. My one unusual bit of notation is A^0 (B^0), which denotes that apparatus 1 (2) is absent. For instance, $\rho(\mu_A \mid \lambda,A,B^0)$ is the probability density that an apparatus set to measure A on a system in state λ occupies microstate μ_A, given that no measurement occurs on particle 2. Similarly, $\rho(\mu_A,\mu_B \mid \lambda,A,B)$ is the joint probability density that apparatuses about to measure A and B on a system in state λ lie in microstates μ_A and μ_B, respectively.

In a hidden-variable theory, the quantum state ϕ is epistemic; a system described by ϕ actually occupies a fully-specified state λ. By "ϕ," we mean $\phi(t=0)$, the quantum state in which the particles were prepared. So, $\rho(\lambda \mid \phi)$ denotes the probability density that a pair of particles prepared in quantum state ϕ (at t=0) occupies ontological state λ at later time t, immediately before measurement.

Depending on the theory, this probability density reflects an in-practice or in-principle uncontrollability of the hidden variables. For instance, in David Bohm's pilot-wave theory, a particle always *has* a definite position, encoded by λ. But uncontrollable fluctuations ensure that identically-prepared particles almost always emerge from the source with slightly different initial trajectories. Bohm's law of motion ensures that these different initial trajectories "fan out" so as to reproduce the spatial distribution of the QM wavefunction.

Probability theory allows us to define

$$p(A=a \mid \lambda,B^0) \equiv \int p(A=a \mid \lambda,\mu_A,B^0)\cdot\rho(\mu_A \mid \lambda,A,B^0)\cdot d\mu_A$$

$$p(A=a,B=b \mid \lambda) \equiv \iint p(A=a,B=b \mid \lambda,\mu_A,\mu_B)\cdot\rho(\mu_A,\mu_B \mid \lambda,A,B)\cdot d\mu_A d\mu_B,$$

where the integrals range over all "contributing" apparatus microstates, namely microstates for which $\rho(\mu_A \mid \lambda,A,B^0)>0$ or $\rho(\mu_A,\mu_B \mid \lambda,A,B)>0$.[8] Physically, $p(A=a \mid \lambda,B^0)$ is the 'μ-averaged' probability that a system in state λ would yield A=a upon measurement.

As a convenient shorthand, define

$$p(A=a \mid \lambda,\mu_A,\mu_B) \equiv \Sigma_i p(A=a,B=b_i \mid \lambda,\mu_A,\mu_B),$$

---

[8]Of course, if the apparatus microstates are "discrete" instead of continuous, then $\rho(\mu_A \mid$ ...) becomes a probability, and we sum instead of integrate over the microstates.

where $\{b_i\}$ are the possible measurement outcomes for B. Physically, $p(A=a \mid \lambda,\mu_A,\mu_B)$ is the probability that A-measurement of particle 1 (with apparatus in microstate μ_A) accompanied by B-measurement of particle 2 (with apparatus in microstate μ_B) would yield A=a for particle 1.

According to a hidden-variable theory, the probability that a system prepared in quantum state ϕ would yield a given measurement outcome is found by averaging over the λ states underlying ϕ:

$$p(A=a \mid \phi,B^0) \equiv \int p(A=a \mid \lambda,B^0)\cdot\rho(\lambda \mid \phi,A,B^0)\cdot d\lambda,$$

$$p(A=a,B=b \mid \phi) \equiv \int p(A=a,B=b \mid \lambda)\cdot\rho(\lambda \mid \phi,A,B)\cdot d\lambda.$$

In this notation, $p(A=a \mid \phi,B^0)$ is the probability *according to the hidden-variable theory* that a system prepared in quantum state ϕ would yield A=a upon measurement. If the hidden-variable theory does not reproduce QM's statistical predictions, then $p(A=a \mid \phi,B^0)$ might not equal $P_{QM}(A=a \mid \phi)$, the probability *according to QM* that a system prepared in state ϕ would yield A=a.

In summary, I've defined three levels of measurement-result probabilities. The fundamental probabilities of the form $p(A=a \mid \lambda,\mu_A,B^0)$ depend on the particles' state and also on the measuring apparatus's microstate. By averaging over apparatus microstates, we obtain probabilities of the form $p(A=a \mid \lambda,B^0)$, which specify the likelihood that a system in state λ would yield A=a upon measurement. Of course, if measurement results do not depend on apparatus microstates, then $p(A=a \mid \lambda,B^0)$ is a "fundamental" probability. Finally, we can average over the λ states underlying the quantum state to

obtain $p(A=a \mid \phi,B^0)$. According to the hidden variable theory, this λ-averaged probability predicts the statistics we would "observe" by measuring A on many systems prepared in quantum state ϕ, assuming no B-measurements occur.

§2.5.2. Stochastic locality conditions

"Stochastic Bell locality" is the requirement that, if events *a* and *b* are spacelike separated, then the occurrence of *b* cannot depend directly on *a*. Philosophers usually formalize this locality intuition in terms of probabilities: the objective probability that *b* occurs cannot depend on whether *a* occurs. Therefore, if *b* is correlated with *a*, some screening-off "common cause" must account for the correlation. A correlation between *a* and *b* does not violate stochastic Bell locality if there exists a common cause *c* such that $p(b|a,c)=p(b|c)$, because this equality shows that *a* does not affect the probability of *b*'s occurrence. Rather, the probability that *b* occurs is "set" by *c*. If no such common cause exists, however, then a correlation between *a* and *b* suggests a direct connection between those events, in violation of stochastic Bell locality.

Stochastic Bell locality motivates the specific conditions needed to derive a Bell inequality in a contextual, stochastic framework. Many authors consider Factorizability, the conjunction of Jarrett's Locality and Completeness, to be the important assumption. The locality conditions about the distributions of hidden-variable states are considered auxiliary and discussed less fully. Since these distributional locality assumptions are nontrivial, however, we must examine their physical content in detail. But first, I'll review Factorizability.

"Locality" & Completeness $\Leftrightarrow$ *Factorizability.* Jarrett (1984) discusses

Jarrett Locality:

$$p(A=a \mid \lambda,\mu_A,B^0) = p(A=a \mid \lambda,\mu_A,\mu_B)$$

Jarrett Completeness:

$$p(A=a,B=b \mid \lambda,\mu_A,\mu_B) = p(A=a \mid \lambda,\mu_A,\mu_B)\cdot p(B=b \mid \lambda,\mu_A,\mu_B)$$

Jarrett Locality demands that a measurement-result probability not depend on the setting or microstate of a distant apparatus. If Jarrett Locality fails, then either the apparatuses "conspired" ahead of time to bring about certain correlations, or changing the state of apparatus 2 can instantaneously affect the probabilities associated with particle 1. I'll assume conspiracies don't happen. As Jarrett shows, if the hidden-variable states were sufficiently controllable, then Jarrett Locality violation would allow experimenters to communicate superluminally. For this and related reasons, Jarrett Locality violation indicates an instantaneous causal connection between the two wings of the experiment, under most notions of causality. (See chapter 5 for more discussion of causation.) QM obeys Jarrett Locality, while Bohm's theory violates that condition.

Jarrett Completeness is often written

$$p(A=a \mid \lambda,\mu_A,\mu_B) = p(A=a \mid \lambda,\mu_A,\mu_B,B=b),$$

which is equivalent to the above for nonzero $p(A=a,B=b \mid \lambda,\mu_A,\mu_B)$. This condition requires that a measurement-result probability depend only on the pre-measurement ontological state of the particles and apparatuses, not on the result of a spacelike

separated measurement. Put another way, the particle and apparatus states must serve as the Reichenbachian (1956) common cause of the correlated measurement results; those states must "screen off" measurement results on particle 1 from measurement results on particle 2.

Why should Completeness hold? After all, we expect that measuring particle 1 might provide previously unknown information about particle 2, thereby changing our epistemic (subjective) measurement-result probabilities associated with particle 2. But the objective probabilities considered here reflect the actual state of the particles, not our state of knowledge. A change in the objective probabilities associated with particle 2 *is* a real physical change in that particle's properties. We don't intuitively expect that obtaining a measurement result on particle 1 can nonlocally "influence" the characteristics of particle 2. Completeness rules out precisely this kind of influence. According to Completeness, obtaining a measurement result on particle 1 can tell us something we didn't know about particle 2, but cannot instantaneously alter the objective probabilities associated with particle 2.

Given this, I can now explain my insistence that λ denote the particles' state immediately before measurement, not the particles' state at the source. Completeness is physically unmotivated if written in terms of the source state. An example will illustrate why. Imagine a toy theory in which particles emerging from the source in state λ_0 stochastically evolve into state λ_1 50% of the time and into state λ_2 50% of the time. Suppose that particle 1 makes its "choice," and then subluminally communicates its choice to particle 2. When the particle pair occupies λ_1, it always yields measurement results A=+1 for particle 1 and B=-1 for particle 2. When the particle pair occupies λ_2, it

always yields A=-1 and B=+1. In this theory, Completeness formulated in terms of λ_0 fails, because $p(B=+1 \mid \lambda_0)=.5$ while $p(B=+1 \mid \lambda_0, A=-1)=1$. But this failure does *not* indicate a nonlocal influence of the particle-1 measurement result on particle 2. Obtaining A=-1 does not physically *affect* particle 2, but simply *reveals* that the particles evolved into λ_2 instead of λ_1. For this reason, stochastic Bell locality does not motivate "λ_0-Completeness." By contrast, stochastic Bell locality motivates Jarrett Completeness, which is formulated in terms of λ immediately before measurement:

$$p(B=+1 \mid \lambda_1) = p(B=+1 \mid \lambda_1, A=-1),$$

etc., which the toy theory obeys.

In summary, stochastic Bell locality motivates Completeness only if formulated in terms of λ immediately before measurement, not in terms of λ at the source.

Completeness violation may challenge the "spirit" of relativity, but does not allow the possibility of superluminal signaling unless accompanied by Locality violation; see Jones and Clifton. QM violates Completeness. The classic example is two electrons in their singlet state. But in fact, QM violates completeness for *any* entangled state.

Deterministic theories are necessarily Complete, but not vice versa (cf. Elby 1990).

The logical conjunction of Jarrett Locality and Jarrett Completeness is equivalent to Factorizability:

Factorizability:

$$p(A=a,B=b \mid \lambda,\mu_A,\mu_B) = p(A=a \mid \lambda,\mu_A,B^0)\cdot p(B=b \mid \lambda,\mu_B,A^0).$$

Throughout this chapter, I'll often invoke Factorizability instead of separately invoking Jarrett Locality and Jarrett Completeness.

Distributional Locality Assumptions: Particle Locality. To derive a Bell inequality, we must make locality assumptions about (i) the distribution of particle hidden-variable states, and (ii) the distribution of apparatus hidden-variable states. The first of these is

Particle Locality:

$$\rho(\lambda \mid \phi,A,B) = \rho(\lambda \mid \phi,A,B^0) = \rho(\lambda \mid \phi,A^0,B^0)$$

where the settings are chosen after the particles have left their source, but immediately before the particles start to interact locally with the apparatuses (i.e., immediately before the measurements begin). This condition formalizes the assumption used above in subsection 2.4.2.

To make Particle Locality more precise, suppose that preparation of the apparatuses occurs during a short time interval Δt, which begins at t_1 and ends at t_2. Suppose further that no non-superluminal "signal" emitted from the devices during Δt could reach the particles until after t_2. In other words, measurement--by which I mean the local interaction between the particles and the apparatuses--cannot begin until after t_2. Particle Locality demands that the probability density for the particles to occupy state λ at time t_2 be the same as if one or both apparatus preparations had never taken place.

Stochastic Bell locality clearly motivates Particle Locality. Particle Locality requires the particles' state not to be instantaneously disturbed by an event, specifically the preparation of a measuring device, that occurs spacelike separated from the particles. Therefore, failure of Particle Locality indicates either a pre-planned conspiracy or a superluminal link between the particles and their measuring devices.

We now see that the physical content of Particle Locality nearly duplicates the physical content of Jarrett Locality. Put another way, if some physical mechanism mediates Jarrett Locality violation, then we intuitively expect the same mechanism to be capable of mediating Particle Locality violation, and vice versa. For this reason, theories in which Jarrett Locality holds while Particle Locality fails are no less plausible than theories in which Particle Locality holds while Jarrett Locality fails. Consequently, we shouldn't consider Jarrett Locality to be "primary" and Particle Locality to be "auxiliary." Instead, we should place these two conditions on equal footing.

To illustrate this point, consider Bohm's theory, in which the quantum wavefunction $\phi(x,t)$ is a "pilot wave" that guides a particle's position (as encoded by λ). The particle and its wavefunction are separately "real" physical things. The wavefunction evolves according to Schrödinger's equation.

In this theory, Jarrett Locality fails. If the two-particle wavefunction is entangled, then measuring particle 1 causes the two-particle wavefunction to entangle with the wavefunction of apparatus 1. This "disturbance" of the wavefunction, the precise nature of which depends on the setting of apparatus 1, instantaneously alters the trajectory of particle 2, thereby altering certain measurement results on particle 2. So, by changing the setting of apparatus 1, you can alter the trajectory (and measurement results) of

particle 2. (As noted above, however, since the particle trajectories are unknown and uncontrollable, you can't use this nonlocality to signal superluminally.)

Does this mechanism of nonlocal wavefunction entanglement in Bohm's theory also lead to violation of Particle Locality? Yes! To see why, consider a single particle approaching Stern-Gerlach magnets. The particle has a definite trajectory, and the statistical spatial distribution of identically-prepared particles is determined by the wavefunction $\phi(x,t)$. Formally,

$$\rho(\lambda \mid \phi)=|\phi(x,t)|^2,$$

where λ is the state corresponding to particle position x at time t. If the particle's state is entangled, replace $|\phi(x,t)|^2$ with the relevant density matrix element. In general, a wavefunction has long "tails" that extend in front of, and behind, the particle "carried" by that wavefunction. Consequently, part of the particle's wavefunction impinges upon the measuring apparatus before the particle itself reaches the apparatus. When this happens, the particle's wavefunction interacts with the apparatus's wavefunction, leading to an entangled wavefunction. This entangled wavefunction *instantaneously* starts to guide the particle (and the apparatus); and the particle's "new" trajectory in general differs from what it would have been had the apparatus been absent, or had the apparatus (i.e., the magnets) been "set" differently (i.e., tilted at a different angle). At a statistical level, the distribution of particle trajectories in the presence of a Stern-Gerlach apparatus set to measure S_x differs from the distribution of particle trajectories in the presence of a Stern-Gerlach apparatus set to measure S_z. Crucially, the differences in trajectories kick in

before the apparatus has time to interact locally with the particle, due to the instantaneous entanglement between the particle's spread-out wavefunction and the apparatus's wavefunction. So, Particle Locality fails.

Let me repeat the argument of the previous paragraph less rigorously but more intuitively. The spatial orientation of the Stern-Gerlach magnets contributes to the potential (or if you prefer, boundary conditions) in which the particle's wavefunction evolves. Consequently, when the particle's spatial wavefunction impinges upon the apparatus, the wavefunction begins evolving differently than it otherwise would have. This change in the wavefunction's spatial evolution instantaneously affects the whole wavefunction, not just the "part" of the wavefunction near the apparatus. So, the presence of the apparatus affects the particle's wavefunction even before the particle itself reaches the apparatus. Furthermore, the alignment of the magnets (i.e., the "setting") determines the shape of the potential in which the particle's wavefunction evolves. In brief, the particle wavefunction's evolution is instantaneously affected by the apparatus setting. This instantaneous change in the particle's wavefunction immediately affects the particle's trajectory. So, Particle Locality fails. Keep in mind, though, that the "real" reason Particle Locality fails in Bohm's theory is wavefunction entanglement between the particle and apparatus.

Here's the point. In Bohm's theory, Jarrett Locality fails because nonlocal entanglement between the particles and apparatuses instantaneously changes the "pilot wave" guiding the particles, and therefore instantaneously changes the particles' trajectories. Particle Locality fails *for the same reason.* The physical mechanism mediating Jarrett Locality violation also mediates Particle Locality violation. Bohm's

theory confirms the intuition that a physical mechanism behind Locality violation is likely also to cause Particle Locality violation, and vice versa. Particle Locality is not a weak auxiliary assumption that we can safely ignore.

Distributional Locality Assumptions: TAF. Now that I've shown the importance of locality assumptions about the distribution of hidden variables, let's consider the distributional locality assumption concerning apparatus microstates:

Total Apparatus Factorability (TAF):

$$\rho(\mu_A,\mu_B \mid \lambda,A,B) = \rho(\mu_A \mid \lambda,A,B^0)\cdot\rho(\mu_B \mid \lambda,A^0,B).$$

According to TAF, the likelihood that an apparatus occupies a given microstate depends only on the setting of that apparatus (and perhaps on the state of the particle it's about to measure), *not* on the setting or microstate of a distance apparatus. The settings are chosen late enough so that the apparatuses could not "communicate" subluminally before the measurements occur.

TAF encodes two physical intuitions. First, the two measuring devices are ontologically separable, as opposed to holistically entangled, and therefore it makes sense to specify the states of the two apparatuses separately. Second, changing the state of apparatus 1 should not affect the state (or more precisely, the state-occupation probabilities) of apparatus 2. Failure of TAF indicates either a pre-planned conspiracy, or else an instantaneous nonlocal connection (perhaps holistic, perhaps causal) between the two apparatuses. Stochastic Bell locality motivates TAF, just as it motivates Factorizability.

During an EPR-type experiment, Bohm's theory violates TAF. (I omit the proof.) This should come as no surprise, because the physical mechanism behind that violation is wavefunction entanglement--the same holistic entanglement that ultimately leads, in Bohm's theory, to Jarrett Locality violation and Particle Locality violation.

Spectrum Rule. In this probabilistic framework, I'll use a version of the Spectrum that a hidden-variable theory *must* obey in order to reproduce QM's predictions.

Spectrum Rule: $p(A \neq \{\text{one of the eigenvalues of } \mathbf{A}\} \mid \phi,...) = 0.$

Recall that, in my notation, p(... | ...) is a probability according to the hidden-variable theory. In words, the Spectrum Rule requires that no matter what quantum state the system occupies, the probability according to the hidden-variable theory that measurement of an observable yields a non-eigenvalue of the corresponding operator is 0. This does not mean that $p(A \neq \{\text{one of the eigenvalues of } \mathbf{A}\} \mid \lambda,...)=0$ for all λ. It merely means that these "anomalous" λ's constitute a zero-measure subset of all the hidden-variable states underlying the quantum state.

To keep all this straight, let me introduce some terminology. State λ "mirrors the spectrum rule with respect to observable A" if $p(A \neq \{\text{one of the eigenvalues of } \mathbf{A}\} \mid \lambda,B^0)=0$.

Summary. Factorizability, Particle Locality, and TAF are the standard assumptions used to derive a stochastic Bell inequality; see Clifton *et al.* (1991). (In most presentations, TAF is omitted, because apparatus microstates aren't considered.) These conditions, motivated by stochastic Bell locality, encode similar physical content.

Usually, Factorizability (i.e., Jarrett Locality and Completeness) is considered the "primary" assumption, while Particle Locality and TAF are considered auxiliary. As argued above, however, a physical mechanism responsible for Jarrett Locality violation is likely to generate Particle Locality or TAF violation as well; and vice versa. Bohm's theory illustrates this point. We should not think of Particle Locality or TAF as auxiliary conditions, because they encode nontrivial physical content.

§2.5.3. Nonlocality theorem

In this section, I prove a Heywood-Redhead-Brown-Svetlichny style algebraic (perfect correlations) nonlocality theorem from the stochastic locality conditions just discussed. Before getting started, I need to prove a trivial lemma:

Factorizability lemma:

Factorizability & TAF $\rightarrow$ $p(A=a,B=b \mid \lambda) = p(A=a \mid \lambda,B^0)\cdot p(B=b \mid \lambda,A^0)$

Proof:

$p(A=a,B=b \mid \lambda) = \iint p(A=a,B=b \mid \lambda,\mu_A,\mu_B)\cdot\rho(\mu_A,\mu_B \mid \lambda,A,B)\cdot d\mu_A d\mu_B$

by definition

$= \iint p(A=a \mid \lambda,\mu_A,B^0)\cdot p(B=b \mid \lambda,\mu_B,A^0)\cdot\rho(\mu_A,\mu_B \mid \lambda,A,B)\cdot d\mu_A d\mu_B$

by Factorizability

$= \iint p(A=a \mid \lambda,\mu_A,B^0)\cdot p(B=b \mid \lambda,\mu_B,A^0)\cdot\rho(\mu_A \mid \lambda,B^0)\cdot\rho(\mu_B \mid \lambda,A^0)\cdot d\mu_A d\mu_B$

by TAF

$= [\int p(A=a \mid \lambda,\mu_A,B^0)\cdot\rho(\mu_A \mid \lambda,B^0)\cdot d\mu_A]\cdot[\int p(B=b \mid \lambda,\mu_B,A^0)\cdot\rho(\mu_B \mid \lambda,A^0)\cdot d\mu_B]$

separating variables

$= p(A=a \mid \lambda,B^0)\cdot p(B=b \mid \lambda,A^0)$

by definition.

In words, if the A-wing and B-wing probabilities are completely independent (as required by Factorizability and TAF), then this probabilistic independence remains after we average over the apparatus microstates. This "μ-less" version of Factorizability is what I'll invoke in the proof below.

Theorem: Factorizability (i.e., Jarrett Locality and Jarrett Completeness) & Particle Locality & TAF → Contradiction with QM's perfect correlations

Proof: As in Brown and Svetlichny's proof, consider two spin-1 particles in their singlet state,

$$|\Psi_{singlet}\rangle = -3^{-1/2}(|S_x=0\rangle\otimes|S_x=0\rangle - |S_y=0\rangle\otimes|S_y=0\rangle + |S_z=0\rangle\otimes|S_z=0\rangle)$$

On particle 1, we'll measure the spin-Hamiltonian $H_S=aS_x^2+bS_y^2+cS_z^2$, while on particle 2 we'll measure one of corresponding spin components, either S_x, S_y, or S_z. The eigenvalues of $\mathbf{H}_S$ are $\{h_x=b+c, h_y=a+c, h_z=a+b\}$.

In this proof, I'll consider 16x12=192 perfect anticorrelations of the form $P_{QM}(H_S=..., S_n=... \mid \Psi_{singlet})=0$. Sixteen is the number of spin-Hamiltonians I'll invoke, corresponding to the 16 orthogonal triads used in the Peres-Kochen-Specker theorem; and for each spin-Hamiltonian, I consider 12 perfect anticorrelations. These exact numbers aren't important. What's important is that I consider only a finite number of perfect anticorrelations.

Consistency with QM demands that the hidden-variable theory reproduce these 192 perfect anticorrelations and obey the spectrum rule. In appendix 2.5.5 below, I prove that, if Particle Locality holds, then a measure-1 subset of the λ states underlying $\Psi_{singlet}$

(i) mirror *each* of these 192 perfect anticorrelations, and also

(ii) mirror the spectrum rule with respect to each of the observables considered here.

Formally, a λ state "mirrors" a QM perfect anticorrelation $P_{QM}(A=a, B=b \mid \Psi)=0$ if $p(A=a, B=b \mid \lambda)=0$.

For the remainder of this proof, let λ denote any member of this measure-1 subset. Crucially, we need to consider only *one* such λ. This fact acquires greater importance in section 2.7.

Kochen-Specker contradiction. I now show that any λ obeying (i) and (ii) generates a Peres-Kochen-Specker contradiction. The only assumption I'll invoke is μ-less Factorizability, which I showed above (in the Factorizability lemma) to follow from Factorizability and TAF.

Since λ obeys the Spectrum Rule with respect to H_S, the only three H_S-measurement outcomes that can have nonzero probability are h_x, h_y, and h_z. Since probabilities are normalized,

$$p(H_S=h_x \mid \lambda,S^0) + p(H_S=h_y \mid \lambda,S^0) + p(H_S=h_z \mid \lambda,S^0) = 1.$$

("S^0" denotes that no measurement occurs on particle 2.) It follows that at least one of those three spin-Hamiltonian measurement-result probabilities is greater than zero.

- Suppose $p(H_s=h_x \mid \lambda,S^0) > 0$.

From QM, we have the following four perfect anticorrelations:

(a) $P_{QM}(H_s=h_x, S_x=+1 \mid \Psi_{singlet}) = 0,$

(a') $P_{QM}(H_s=h_x, S_x=-1 \mid \Psi_{singlet}) = 0,$

(b) $P_{QM}(H_s=h_x, S_y=0 \mid \Psi_{singlet}) = 0,$

(c) $P_{QM}(H_s=h_x, S_z=0 \mid \Psi_{singlet}) = 0.$

Since λ reproduces these anticorrelations,

(a) $p(H_s=h_x, S_x=+1 \mid \lambda) = 0,$

(a') $p(H_s=h_x, S_x=-1 \mid \lambda) = 0,$

(b) $p(H_s=h_x, S_y=0 \mid \lambda) = 0,$

(c) $p(H_s=h_x, S_z=0 \mid \lambda) = 0.$

Since μ-less Factorizability holds, we get

(a) $p(H_s=h_x \mid \lambda,S^0)\cdot p(S_x=+1 \mid \lambda,H^0) = 0,$

(a') $p(H_s=h_x \mid \lambda,S^0)\cdot p(S_x=-1 \mid \lambda,H^0) = 0,$

(b) $p(H_s=h_x \mid \lambda,S^0)\cdot p(S_y=0 \mid \lambda,H^0) = 0,$

(c) $p(H_s=h_x \mid \lambda,S^0)\cdot p(S_z=0 \mid \lambda,H^0) = 0.$

By supposition, $p(H_s=h_x \mid \lambda,S^0)>0$. Therefore,

(a) $p(S_x=\pm 1 \mid \lambda,H^0) = 0$

(b) $p(S_y=0 \mid \lambda,H^0) = 0$,

(c) $p(S_z=0 \mid \lambda,H^0) = 0$.

Since λ obeys the spectrum rule with respect to S_x, normalization implies

$$p(S_x=-1 \mid \lambda,H^0) + p(S_x=0 \mid \lambda,H^0) + p(S_x=+1 \mid \lambda,H^0) = 1.$$

From this and (a), we immediately get $p(S_x=0 \mid \lambda,H^0) = 1$. In summary, we have

(a) $p(S_x=0 \mid \lambda,H^0) = 1$

(b) $p(S_y=0 \mid \lambda,H^0) = 0$,

(c) $p(S_z=0 \mid \lambda,H^0) = 0$.

This conclusion, for the particles in state λ, followed from μ-less Factorizability and the supposition that $p(H_s=h_x \mid \lambda,S^0) > 0$. If we suppose instead that $p(H_s=h_y \mid \lambda,H^0) > 0$, reasoning similar to the above, with x, y, and z cyclically permuted, yields $p(S_x=0 \mid \lambda,H^0)=0$, $p(S_y=0 \mid \lambda,H^0)=1$, and $p(S_z=0 \mid \lambda,H^0)=0$. Similarly, if we suppose $p(H_s=h_z \mid \lambda,S^0) > 0$, we conclude that $p(S_x=0 \mid \lambda,H^0)=0$, $p(S_y=0 \mid \lambda,H^0)=0$, and $p(S_z=0 \mid \lambda,H^0)=1$.

As noted above, by the spectrum rule, at least one of those three spin-Hamiltonian measurement-result probabilities is greater than 0. Therefore, from the previous paragraph, we see that the three values

$$\{p(S_x=0 \mid \lambda,H^0),\ p(S_y=0 \mid \lambda,H^0),\ p(S_z=0 \mid \lambda,H^0)\}$$

must be such that two of the values equal 0 while the third value equals 1.

Due to the spherical symmetry of the spin singlet state Ψ, the same conclusion applies to *each* of the 16 orthogonal triads of directions needed to generate the Kochen-Specker-Peres contradiction.

As noted above, each point on the unit sphere is associated with a unit vector (i.e., a direction) **n**. For each of the 33 **n**'s contained in the 16 orthogonal triads, map to **n** the value $p(S_n=0 \mid \lambda,H^0)$. As just shown, this map is such that for any orthogonal triad, two points take on the value 0 while the third point takes on the value 1. But such a map is algebraically impossible, by the Kochen-Specker-Peres contradiction. This contradiction establishes that no theory obeying the stochastic Bell locality conditions discussed above can reproduce the perfect anticorrelations of QM. *Q.E.D.*

§2.5.4. Discussion

This was the first algebraic (perfect correlations) nonlocality proof that used stochastic as opposed to deterministic locality assumptions. The "trick" was to associate points on the unit sphere with *probabilities* instead of possessed values. Previously, it had seemed that, to rule out stochastic local hidden-variable theories, it would be

necessary to consider the statistical correlations of QM, as the Bell inequalities do. But actually, something about the algebraic structure of quantum theory--as reflected in the perfect correlations--already rules out stochastic locality. In section 2.6, I'll argue that these perfect correlations reflect underlying conservation principles, and are therefore in a sense more "fundamental" than the general statistical predictions of QM.

Critics could point out that my theorem isn't as "stochastic" as it initially appears. Suppes and Zanotti (1976) prove that any Factorizable theory reproducing all the perfect correlations of QM is necessarily deterministic, in the sense that all the probabilities "collapse" to zero or one.

In response, I can point to sections 2.6 and 2.7 below. In section 2.6, I relax the requirement that the hidden-variable theory *exactly* reproduce QM's perfect correlations. It might turn out that the perfect correlations under discussion are only approximations to a true theory incorporating tiny deviations from the perfect correlations. I prove below that such a theory cannot be Bell local. Since the hidden-variable theories "captured" by that proof do not reproduce the perfect correlations, they escape the Suppes-Zanotti collapse; truly stochastic theories get ruled out by the theorem. Similarly, in section 2.7, I weaken the above stochastic Bell locality assumptions, and prove that any theory obeying even those weakened conditions cannot exactly reproduce the perfect correlations. Since the proof relies on a condition weaker than Factorizability, it escapes the Suppes-Zanotti proof.

So, the main philosophical "work" done by the above proof is to open a new avenue of investigation into algebraic (perfect correlations) nonlocality proofs. Specifically, I showed how to use stochastic locality assumptions directly in such proofs, without

invoking the Suppes-Zanotti collapse. Given this new "tool," we can explore stochastic local theories in contexts where they do and do not collapse into deterministic theories.

§2.5.5. *APPENDIX: Lemma from theorem 2.5.3*

Particle Locality implies that , if the hidden-variable theory reproduces QM's predictions, then a measure-1 subset of the λ states underlying Ψ

(i) mirror each of the 192 perfect anticorrelations used in theorem 2.5.3; and

ii) mirror the Spectrum Rule with respect to each of the 16 spin-Hamiltonians and 33 spin components considered here.

Proof: Suppose the theory reproduces QM's perfect anticorrelations and obeys the Spectrum Rule, as required by consistency with QM. Let $P_{QM}(A=a, B=b \mid \Psi)=0$ denote any one of the 192 perfect anticorrelations invoked in this proof. The corresponding hidden-variable theory probability is

$$\begin{aligned} p(A=a, B=b \mid \Psi) &= \int p(A=a,B=b \mid \lambda)\cdot\rho(\lambda \mid \Psi,A,B)\cdot d\lambda \\ &= {}_{\Lambda}\int p(A=a,B=b \mid \lambda)\cdot\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda, \end{aligned}$$

where I used Particle Locality in the second line, and where Λ denotes the set of λ states for which $\rho(\lambda \mid \Psi,A^0,B^0)>0$. Since this integral must equal zero in order to reproduce the QM perfect anticorrelation, it follows that a measure-1 subset of Λ is such that each λ in the subset mirrors the perfect anticorrelation, i.e., $p(A=a, B=b \mid \lambda)=0$ for each λ in that subset. So, corresponding to each of the 192 perfect anticorrelations is a measure-1 subset of Λ such that each element of the subset mirrors the perfect anticorrelation.

These 192 subsets are not necessarily equivalent. But from measure theory, the intersection of a *finite* number of measure-1 subsets of Λ is itself a measure-1 subset of Λ. Call this "intersection" subset Λ'. Every λ in Λ' mirrors *each* of the 192 perfect anticorrelations.

Similar considerations apply to the Spectrum Rule. Let A denote one of the 16+33=49 observables considered in this proof. Let $\{a_i\}$ denote the eigenvalues of A. For consistency with QM, the hidden variable theory must give

$$p(A\neq\{a_i\} \mid \Psi) = \int_\Lambda p(A\neq\{a_i\} \mid \lambda,B^0)\cdot\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda = 0,$$

where I again used Particle Locality. Since Λ' is a nonzero-measure subset of Λ, it follows that

$$\int_{\Lambda'} p(A\neq\{a_i\} \mid \lambda,B^0)\cdot\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda = 0.$$

Therefore, a measure-1 subset of Λ' is such that each λ in the subset mirrors the Spectrum Rule with respect to A, i.e., $p(A\neq\{a_i\} \mid \lambda,B^0)=0$ for each λ in the subset. Hence, corresponding to each of the 49 observables is a measure-1 subset of Λ' such that each λ in the subset mirrors the Spectrum Rule with respect to that observable. The intersection of these 49 measure-1 subsets of Λ' is itself a measure-1 subset of Λ'. Call this intersection subset Λ''. Each element of Λ'' not only (i) mirrors all 192 perfect anticorrelations, but also (ii) mirrors the Spectrum Rule with respect to all 49 observables used in this proof. And Λ'' is a measure-1 subset of Λ', which is itself a

measure-1 subset of Λ, the hidden-variable states underlying the quantum state. In summary, a measure-1 subset of the λ states underlying Ψ obey (i) and (ii).

To reach this conclusion, I required the hidden-variable theory to reproduce QM's predictions, and to obey Particle Locality. *Q.E.D.*

Notice that the proof would have failed if we weren't considering only a finite number of perfect correlations. That's why Bell (1966) and Kochen and Specker (1967) are improvements upon Gleason (1957): Gleason needs an uncountable infinity of orthogonal triads, whereas Kochen and Specker need only a finite number.

Section 2.6: Imperfect correlations nonlocality proof

In the previous section, I showed that no Bell local theory can precisely reproduce the EPR-type perfect anticorrelations of QM. Now I'll show that no local theory can even *approximate* those perfect anticorrelations. This line of reasoning addresses a common criticism of proofs that rely on the perfect anticorrelations: Due to detector inefficiencies, the perfect anticorrelations cannot be confirmed experimentally. At best, we can confirm that they hold to excellent approximation. Therefore, an empirically adequate hidden-variable theory need not *exactly* reproduce those correlations. The new kind of proof introduced in this section prevents a hidden-variable theorist from using this escape route to try to resurrect local causality. After completing the proof, I'll discuss in more detail the philosophical implications of this kind of proof.

§2.6.1. Near-perfect correlations

I'll begin by formalizing the requirement that a theory nearly, but not precisely, reproduce the QM perfect correlations.

The perfect (anti)correlations considered in algebraic nonlocality theorems emerge from fundamental conservation principles. For instance, consider two spin-1 particles prepared in such a way that their total angular momentum is 0. Then the probability that both particles will yield "up" when their n-component of angular momentum gets measured is $P_{QM}(J_n \otimes I=+1, I \otimes J_n=+1 \mid \Psi_{J=0})=0$. This perfect anticorrelation reflects, and in a sense directly encodes, conservation of angular

momentum. The perfect anticorrelations considered in section 2.5, and reconsidered here, also stem from angular momentum conservation.

We know, however, that some conservation laws are approximate instead of absolute. A good example is charge-conjugation/parity (CP) invariance, originally considered fundamental, but now thought to be violated by 'weak nuclear' interactions. Perhaps angular momentum conservation, like CP invariance, is only approximate. Or perhaps some other small interaction "breaks" the perfect correlations stemming from angular momentum conservation. In either case, the perfect anticorrelations predicted by QM for the spin singlet state could fail. But the failure would be small enough so as to escape easy detection. Therefore, an empirically adequate hidden-variable theory would *almost* reproduce those QM anticorrelations. Formally, if angular momentum conservation fails only minutely, we expect the following Near-Perfect Correlations condition to hold:

Near-Perfect Correlations:

$$P_{QM}(Q=q, R=r \mid \phi) = 0 \;\rightarrow\; p(Q=q, R=r \mid \phi) \leq \delta,$$

where $P_{QM}(Q=q, R=r \mid \phi)=0$ is any QM perfect anticorrelation stemming from conservation of angular momentum. Recall that $P_{QM}(... \mid \phi)$ denotes a probability *according to QM,* while $p(... \mid \phi)$ denotes the corresponding probability *according to the hidden-variable theory,* found by averaging over the hidden-variable states underlying the quantum state. The "nearness parameter" δ encodes how closely the hidden-variable theory reproduces the perfect correlations.

As just noted, a hidden-variable theorist could posit such violations for many reasons besides angular momentum non-conservation; see section 2.6.4. But in some theories, failure of Near-Perfect Correlations indicates that angular momentum conservation fails utterly, and cannot be considered even approximate.

Recall from above that we needed to consider 16x12=192 perfect anticorrelations to complete the Peres-Kochen-Specker style proof. I will prove that for $\delta < \frac{1}{192x9} = \frac{1}{1728}$, Near-Perfect Correlations is inconsistent with stochastic Bell locality. If this δ seems low, keep in mind that Clifton *et al.* (1991) have derived a similar result in which δ=0.2. Rob Clifton and I worked together (by FAX) to develop this style of proof.

§2.6.2. Imperfect correlations nonlocality theorem, part I

In this subsection, I'll prove that for $\delta < \frac{1}{192x9} = \frac{1}{1728}$, the Near-Perfect Correlations condition just introduced implies that the hidden-variable theory obeys a mathematical condition I'll call "Fuzzy Correlations." The proof relies only on pure mathematics (e.g., measure theory) and on Particle Locality. In subsection 2.6.3, I'll show that no theory obeying the usual stochastic Bell locality conditions can satisfy Fuzzy Correlations. These two subsections, taken together, prove that no Bell local theory can obey Near-Perfect Correlations. That is, no local theory can even approximate the perfect correlations of QM.

Streamlined Kochen-Specker type arguments may show that we need fewer than 192 anticorrelations to reach a contradiction. To account for that possibility, let N denote the

minimum number of anticorrelations needed to complete a Kochen-Specker style proof in $\mathbb{H}_3 \otimes \mathbb{H}_3$.

Here's the mathematical condition I'll need. Let $P_{QM}(Q_i=q_i, R_i=r_i \mid \Psi)=0$ denote the i-th perfect anticorrelation used in the proof of section 2.5. So, i ranges from 1 to N. The Q_i's are spin-Hamiltonians, and the R_i's are spin-components. But in this section, it doesn't matter what perfect correlations I'm talking about.

Fuzzy Correlations: For a nonzero-measure set of λ states underlying Ψ, $p(Q_i=q_i, R_i=r_i \mid \lambda) < 1/9$ for all i from 1 to N.

In words, Fuzzy Correlations demands that at least *some* of the hidden-variable states underlying Ψ approximately reproduce *all* of the relevant quantum anticorrelations.

Theorem: For $\delta < \frac{1}{9N}$, Near-Perfect Correlations & Particle Locality $\rightarrow$ Fuzzy Correlations.

Proof: By contradiction. Suppose Fuzzy Correlations fails. Then, for each λ belonging to a measure-1 subset of the hidden-variable states underlying Ψ, there exists an i such that $p(Q_i=q_i, R_i=r_i \mid \lambda) \geq 1/9$. In other words, each λ in that measure-1 subset belongs to at least one set $\{\lambda_i\}$, where $\{\lambda_i\}$ denotes the set of states for which $p(Q_i=q_i, R_i=r_i \mid \lambda) \geq 1/9$. Therefore, measure theory trivially implies

$$(*) \qquad m\{\lambda_1\} + m\{\lambda_2\} + m\{\lambda_3\} + \ldots + m\{\lambda_N\} \geq 1,$$

where $m\{\lambda_i\}$ is the measure of set $\{\lambda_i\}$ with respect to $\rho(\lambda \mid \Psi)$.[9]

From (*), it follows that, for at least one i, $m\{\lambda_i\} \geq 1/N$. Consider that i. Recall that for each member of $\{\lambda_i\}$, $p(Q_i=q_i, R_i=r_i \mid \lambda) \geq 1/9$ even though $P_{QM}(Q_i=q_i, R_i=r_i \mid \Psi)=0$. So, for that i, the relevant "observable" probability, according to the hidden-variable theory, is

$$
\begin{aligned}
p(Q_i=q_i, R_i=r_i \mid \Psi) &= \int d\lambda \cdot p(Q_i=q_i, R_i=r_i \mid \lambda) \cdot \rho(\lambda \mid \Psi) \\
&\geq {}_{\{\lambda_i\}}\int d\lambda \cdot p(Q_i=q_i, R_i=r_i \mid \lambda) \cdot \rho(\lambda \mid \Psi) \\
&\qquad \text{since } \{\lambda_i\} \text{ is a subset of the } \lambda\text{'s underlying } \Psi \\
&\geq {}_{\{\lambda_i\}}\int d\lambda \cdot \frac{1}{9} \cdot \rho(\lambda \mid \Psi) \\
&\qquad \text{since } p(Q_i=q_i, R_i=r_i \mid \lambda) \geq 1/9 \text{ for all } \lambda\text{'s in } \{\lambda_i\} \\
&= \frac{1}{9}\,{}_{\{\lambda_i\}}\int d\lambda \cdot \rho(\lambda \mid \Psi) \\
&= \frac{1}{9} m\{\lambda_i\} \\
&\qquad \text{by definition of this measure}
\end{aligned}
$$

---

[9]Formally, $m\{\lambda_i\} \equiv {}_{\{\lambda_i\}}\int \rho(\lambda \mid \Psi) \cdot d\lambda$, where as indicated the integral ranges only over states in $\{\lambda_i\}$. Notice that this measure is uniquely defined only because I've assumed Particle Locality, which implies that $\rho(\lambda \mid \Psi, Q_i, R_i)=\rho(\lambda \mid \Psi, Q^0, R^0)$ for all i. Since the distribution of λ states underlying the quantum state doesn't depend on what's being measured, we can call that distribution $\rho(\lambda \mid \Psi)$.

By the way $\Sigma_i m\{\lambda_i\}$ might be greater than 1, instead of merely equal to 1, since some λ's might belong to more than one set $\{\lambda_i\}$.

$$\geq \frac{1}{9}(\frac{1}{N})$$

since $m\{\lambda_i\} \geq 1/N$ for the i under consideration.

In brief, for the particular perfect anticorrelation under consideration, the hidden-variable theory predicts that $p(Q_i=q_i, R_i=r_i \mid \Psi) \geq \frac{1}{9N}$. But according to Near-Perfect Correlations (with $\delta<\frac{1}{9N}$), $p(Q_i=q_i, R_i=r_i \mid \Psi) < \frac{1}{9N}$. This completes the proof by contradiction. *Q.E.D.*

I just showed that if a theory approximately reproduces the perfect anticorrelations of QM and also obeys Particle Locality, then it necessarily violates a mathematical condition, Fuzzy Correlations.

§2.6.3. Imperfect correlations nonlocality theorem, part II

In this section, I'll prove that no stochastic Bell local theory can obey Fuzzy Correlations. When combined with the theorem just proven, this result shows that no local theory can obey Near-Perfect Correlations.

Theorem: Factorizability & TAF & Spectrum Rule & Fuzzy Correlations $\rightarrow$ Kochen-Specker contradiction.

Proof:

Let λ be a state such that the implication

$$P_{QM}(Q_i=q_i, R_i=r_i \mid \Psi) = 0 \rightarrow p(Q_i=q_i, R_i=r_i \mid \lambda) < 1/9$$

holds for each of the N=192 anticorrelations considered in this proof. By Fuzzy Correlations, a nonzero-measure set of λ states satisfies this condition.

We'll now focus on 12 of these 192 perfect anticorrelations, namely those involving $H_s=aS_x^2+bS_y^2+cS_z^2$ for a particular orthogonal triad of directions, {**x**,**y**,**z**}.

- Suppose $p(H_s=h_x \mid \lambda) \geq 1/3$ on particle 1, when particle 2 isn't measured. From QM, we have the following four perfect anticorrelations:

(a) $P_{QM}(H_s=h_x, S_x=+1 \mid \Psi) = 0,$

(a') $P_{QM}(H_s=h_x, S_x=-1 \mid \Psi) = 0,$

(b) $P_{QM}(H_s=h_x, S_y=0 \mid \Psi) = 0,$

(c) $P_{QM}(H_s=h_x, S_z=0 \mid \Psi) = 0.$

Fuzzy Correlations, applied to those four equalities, implies

(a) $p(H_s=h_x, S_x=+1 \mid \lambda) < 1/9,$

(a') $p(H_s=h_x, S_x=-1 \mid \lambda) < 1/9,$

(b) $p(H_s=h_x, S_y=0 \mid \lambda) < 1/9,$

(c) $p(H_s=h_x, S_z=0 \mid \lambda) < 1/9.$

Applying μ-less Factorizability (implied by Factorizability and TAF) yields

(a) $p(H_s=h_x \mid \lambda) \cdot p(S_x=+1 \mid \lambda) < 1/9,$

(a') $p(H_s=h_x \mid \lambda)\cdot p(S_x=-1 \mid \lambda) < 1/9$,

(b) $p(H_s=h_x \mid \lambda)\cdot p(S_y=0 \mid \lambda) < 1/9$,

(c) $p(H_s=h_x \mid \lambda)\cdot p(S_z=0 \mid \lambda) < 1/9$.

By supposition, $p(H_s=h_x \mid \lambda) \geq 1/3$. From simple algebra, it follows that

(a) $p(S_x=+1 \mid \lambda) < 1/3$,

(a') $p(S_x=-1 \mid \lambda) < 1/3$,

(b) $p(S_y=0 \mid \lambda) < 1/3$,

(c) $p(S_z=0 \mid \lambda) < 1/3$.

According to the Spectrum Rule, the only three outcomes of measuring S_x that have nonzero probability are {-1,0,+1}. By normalization we then have

$$p(S_x=-1 \mid \lambda) + p(S_x=0 \mid \lambda) + p(S_x=+1 \mid \lambda) = 1.$$

From this and inequalities (a) and (a'), we get $p(S_x=0 \mid \lambda) > 1 - 2/3$; that is, $p(S_x=0 \mid \lambda) > 1/3$.

In summary, we have

(a) $p(S_x=0 \mid \lambda) > 1/3$,

(b) $p(S_y=0 \mid \lambda) < 1/3$,

(c) $p(S_z=0 \mid \lambda) < 1/3$.

Now I define a mathematical step function, which has no physical interpretation or importance:

$$K(x) = 0 \text{ if } x < 1/3$$
$$1 \text{ if } x \geq 1/3.$$

Applying the K function to inequalities (a), (b), and (c) yields

(a) $K(p(S_x=0 \mid \lambda)) = 1$,

(b) $K(p(S_y=0 \mid \lambda)) = 0$,

(c) $K(p(S_z=0 \mid \lambda)) = 0$.

This conclusion followed from Factorizability, TAF, Fuzzy Correlations, Spectrum Rule, and the supposition that $p(H_s=h_x \mid \lambda) \geq 1/3$. If we suppose instead that $p(H_s=h_y \mid \lambda) \geq 1/3$, similar reasoning (cyclically permuting **x**, **y**, and **z**) yields $K(p(S_x=0 \mid \lambda)) = 0$, $K(p(S_y=0 \mid \lambda)) = 1$, and $K(p(S_z=0 \mid \lambda)) = 0$. Or, if we suppose that $p(H_s=h_z \mid \lambda) \geq 1/3$, we get $K(p(S_x=0 \mid \lambda)) = 0$, $K(p(S_y=0 \mid \lambda)) = 0$, and $K(p(S_z=0 \mid \lambda)) = 1$.

In summary, if $p(H_s=h_x \mid \lambda) \geq 1/3$, *or* if $p(H_s=h_y \mid \lambda) \geq 1/3$, *or* if $p(H_s=h_z \mid \lambda) \geq 1/3$, then the three values

$$\{ K(p(S_x=0 \mid \lambda)),\ K(p(S_y=0 \mid \lambda)),\ K(p(S_z=0 \mid \lambda)) \}$$

are such that two of the values equal 0 while the third value equals 1. But by normalization and the Spectrum Rule,

$$p(H_s=h_x \mid \lambda) + p(H_s=h_y \mid \lambda) + p(H_s=h_z \mid \lambda) = 1,$$

from which it follows that at least one of those three spin-Hamiltonian measurement-result probabilities is greater than or equal to 1/3. Therefore, the three values

$$\{ K(p(S_x=0 \mid \lambda)),\ K(p(S_y=0 \mid \lambda)),\ K(p(S_z=0 \mid \lambda)) \}$$

are indeed such that two of the values equal 0 while the third value equals 1.

Due to the spherical symmetry of the quantum spin singlet state, the same argument applies to all orthogonal triads of directions {**x**,**y**,**z**} corresponding to the spin-Hamiltonian and spin-component operators utilized in this style of Kochen-Specker proof. (See section 2.5 above.) So, by mapping the value $K(p(S_n=0 \mid \lambda))$ to each unit vector **n** considered in the Kochen-Specker-Peres theorem, we generate an inconsistent map. *Q.E.D.*

This completes the proof that Fuzzy Correlations contradicts either the stochastic Bell locality conditions (Factorizability, TAF, and Particle Locality) or the Spectrum Rule (a violation of which would immediately contradict the predictions of QM).

§2.6.4. "Orthodox spin" theories and conservation

In section 2.6.2, I proved that any theory that approximately reproduces a particular set of QM perfect anticorrelations must obey a mathematical condition called Fuzzy Correlations. Then, in section 2.6.3, I proved that Fuzzy Correlations contradicts the stochastic Bell locality conditions (assuming the Spectrum Rule). So, no Bell-local theory obeying the Spectrum Rule can even approximately reproduce the EPR-type perfect correlations of QM. In this section, I'll examine the philosophical implications of this result by focusing on the connection between perfect correlations and conservation principles.

As discussed in above, the quantum mechanical perfect anticorrelations invoked in my theorems emerge from conservation of angular momentum, which in turn follows from rotational invariance. In some hidden-variable theories, those anticorrelations also emerge from rotational invariance. Call such constructions "orthodox spin" theories.

Orthodox spin theory: Let T denote all the first principles of a theory other than rotational invariance. The theory is an orthodox spin theory iff (a) Spin 'observables' obey the Spectrum rule, and (b) T & (rotational invariance) $\rightarrow$ (the perfect anticorrelations used in the above theorems).

This definition does not presuppose that rotational invariance is a postulate of an orthodox spin theory, but *does* suppose rotational invariance to be consistent with T.

In some theories, of course, rotational invariance doesn't appear as a separate first principle, but instead gets "built into" other postulates. If such a theory obeys the

Spectrum Rule (for spin observables) and reproduces the relevant perfect anticorrelations, then it's an orthodox spin theory.

We now have

Theorem: An orthodox spin theory either violates stochastic Bell locality or violates relativity,

which follows trivially from theorems 2.6.2 and 2.6.3, the definition of orthodox spin theories, and the fact that rotational invariance is a first principle of relativity.

Later, I'll discuss the dilemma this theorem poses for orthodox spin theorists. But first, I explore which theories fit that description.

Any theory obeying the following three conditions is an orthodox spin theory:

(A) Some particles display a discretized intrinsic ('spin') angular momentum. Measurement of a spin component yields ± 1 or 0, in appropriate units.

(B) For those particles, the perfect anticorrelations invoked above follow, in part, from conservation of angular momentum.

(C) Conservation of angular momentum follows from rotational invariance.

Condition (A) receives strong, though indirect support from Stern-Gerlach type experiments, in which a beam of spin-1 particles gets split into three beams upon passing between the magnets. To claim that those experiments support (A), we must assume a certain relation between a particle's spin and magnetic moment. (The copious

"direct" evidence from particle accelerator experiments that spin-1 particles exist is "evidence" only if we assume conservation of angular momentum--an assumption we can't make lightly in the present discussion!)

Condition (C) holds not only for quantum mechanics and quantum field theory, but also for classical mechanics, classical electrodynamics, and special relativity. Noether's theorem shows that whenever equations of motion can be derived via variational calculus from a Lagrangian, symmetries of the Lagrangian lead to conserved quantities. I know of no present theory in which rotational invariance doesn't lead to a conserved "angular momentum" quantity.

Condition (B) is perhaps the fishiest. In quantum mechanics, (B) holds because spin is quantized and particles exist in "superposition" states of indefinite n-component of angular momentum, among other reasons. A general theory might not incorporate all these features, and hence (B) could fail. (B) could also fail because an undetected form of angular momentum or of spin-orbit coupling exists. Yet, (B) may hold for a large class of theories that propose small corrections to quantum mechanics without overhauling the whole theory.

(A), (B), and (C) are sufficient conditions for an orthodox spin theory. But they aren't necessary. Notably, a "spinless" version of David Bohm's (1987) construction violates (B) *and* (A) but is nonetheless an "orthodox spin" theory. In this theory, particles don't have intrinsic angular momentum. Stern-Gerlach experiments turn out the way they do because of an elaborate interaction, mediated by a 'quantum potential,' between the particle and the magnet. Nonetheless, in Bohm's theory, the EPR perfect

anticorrelations emerge from rotational invariance of the relevant quantum potential. Indeed, all hidden-variable theories with which I'm familiar are orthodox spin theories.

In summary: Although we have limited *a priori* motivation for singling out orthodox spin theories, such theories are plausible and important. Quantum mechanics itself, along with the best-developed hidden-variable constructions, are orthodox spin theories. Therefore, no-go results about orthodox spin theories deserve philosophical analysis.

In the next subsection, I show that orthodox spin theorists must renounce at least the spirit of relativity.

§2.6.5. Locality and the spirit of relativity

As theorem 2.6.4 shows, an orthodox spin theorist must abandon either rotational invariance or Bell locality. Failure of stochastic Bell locality violates at least the *spirit* of relativity theory, as I now argue.

In my view, the spirit of relativity demands that the physical characteristics of a system (and its measuring device) be affected *only* by events or states-of-affairs in the backward light cone of that system (and measuring device). Therefore, by the spirit of relativity, neither putting the particle-2 measuring device into a certain state, nor obtaining a measurement result on particle 2, may instantaneously affect the ontological measurement-result probabilities associated with particle 1 and its measuring apparatus. Sure, measuring particle 2 may change our state of knowledge about particle 1, by revealing previously-unknown information. But measuring particle 2 may not change the physical properties of particle 1.

Assuming no "conspiracies," failure of Locality, Particle Locality, or TAF almost certainly constitutes a nonlocal causal link (under most notions of causality), in violation of the spirit of relativity. And recall from section 2.5.2 that when Completeness fails, obtaining a measurement outcome on particle 2 actually *changes* the propensities of particle 1, instantaneously at a distance. Therefore, violation of Completeness, though consistent with the relativistic formalism, constitutes a nonlocal connection that also violates the spirit of relativity as just defined. This conclusion holds no matter whether you consider the nonlocality to stem from a "causal" link or from a "holistic" connection between the particles. (I'll address the causality vs. holism issue in chapter 5.)

So, theorem 2.6.4 raises a dilemma for orthodox spin theorists. Either they must abandon stochastic Bell locality, thereby violating the spirit of relativity; or they must abandon rotational invariance, thereby contradicting the formalism of relativity. The irony is this: Even though Bell locality encodes the spirit of relativity, Bell locality is logically inconsistent with relativity for orthodox spin theories.

Clifton *et al.* (1991), working along different lines, have also derived an "imperfect correlations" algebraic proof. Their work can be used to show that Near-Perfect Correlations contradicts stochastic Bell locality, though they do not do so explicitly. The advantage of their proof, which does not invoke the Kochen-Specker contradiction, is its reliance on a very small number of anticorrelations. As a result, the δ Clifton *et al.* would get in their Near-Perfect Correlations condition is δ=0.2, about 350 times larger than mine.

Clifton *et al. stress* the experimental implications of their imperfect correlations proof. Specifically, they believe the predictions of QM are correct, so that experimental

deviations from the perfect correlations stem from detector inefficiencies. Only an ideal detector could confirm QM's perfect correlations. But an imperfect detector can verify Near-Perfect Correlations for large enough δ. Therefore, Clifton's work allows a practical perfect-correlations experiment to rule out stochastic Bell locality. Furthermore, Clifton *et al.* note, their experiment could improve slightly on Bell-type experiments by showing that a higher fraction (i.e. measure) of λ states contradict one of the stochastic Bell locality conditions. See Clifton *et al.* (1991) for details.

My focus, on the other hand, is more abstractly philosophical. Independent of whether an experiment can in practice verify my Near-Perfect Correlations assumption (with $\delta<1/1728$), I'm interested in the dilemma raised by the *logical* contradiction between Near-Perfect Correlations and stochastic Bell locality (assuming the Spectrum Rule). This contradiction forces a local realist to deny that certain quantum correlations hold even *approximately*. And this contradiction forces an orthodox spin theorist to renounce either the spirit of relativity theory as encoded by Bell locality, or relativity theory itself.

§2.6.6. Summary

In this section, I took advantage of my new framework for proving algebraic (perfect-correlations) nonlocality proofs using probabilities instead of possessed values. Specifically, working within this probabilistic framework, I showed that a stochastic Bell local theory cannot even approximate the perfect correlations of QM, correlations that stem from fundamental conservation principles. This result not only demonstrates the power of working within a stochastic framework, but also helps to quash the hopes of

"local realist" hidden-variable theorists who hope to circumvent nonlocality no-go theorems by proposing small "corrections" to QM or by proposing small but essential detector inefficiencies. Furthermore, this result underscores the tension between the spirit of relativity theory (as encoded by Bell locality) and the letter of relativity theory (specifically, rotational invariance), in quantum framework.

Section 2.7: Imperfect correlations nonlocality proof

Here, I'll modify the main theorem of section 2.5 in order to derive an algebraic (perfect-correlations) nonlocality theorem from assumptions *weaker* than the usual stochastic Bell locality conditions (Factorizability, TAF, and Particle Locality). To my knowledge, no nonlocality result uses weaker assumptions.[10] I'm working on this project with Martin Jones.

The proof involves some tedious measure-theoretic reasoning that I've relegated to an appendix (section 2.7.6). After deriving the relevant technical result, I'll discuss the philosophical implications. In a nutshell, here's the scoop: The standard stochastic Bell locality conditions encode the requirement that the occurrence of an event not affect the *probability* of a spacelike separated event. By contrast, our weakened locality conditions allow event *a* to affect the probability of spacelike separated event *b*. The weakened locality conditions demand only that, roughly speaking, event *a* not affect the *possibility* of event *b* (i.e., *a* may not affect whether or not *b* is possible). More on this later. First, we've got some technical results to wade through.

§2.7.1. Weakened locality assumptions

First, I'll introduce the three weakened locality conditions, briefly discussing their physical content. Then, in section 2.7.2, I'll prove that these weakened conditions contradict the QM perfect anticorrelations invoked above.

---

[10]Remember, derivations relying on counterfactual definiteness implicitly assume determinism, which is stronger than Completeness. Stapp (1993, 1994) proves a nonlocality theorem involving counterfactuals *without* counterfactual definiteness, under certain versions of modal logic. Stapp's locality conditions are probably neither stronger nor weaker than mine. Explicating the precise logical and physical relationships between Stapp's locality assumptions and "standard" locality assumptions constitutes an interesting but difficult project which I won't undertake here.

Each of the three standard stochastic Bell locality assumptions--Factorizability, Particle Locality, and TAF--can be weakened. In these conditions, A (B) refers to an observable associated with particle 1 (2). A^0 (B^0) denotes a *lack* of a measurement performed on particle 1 (2).

Weak Factorizability:

$$p(A=a \mid \lambda,\mu_A,B^0) > 0 \text{ and } p(B=b \mid \lambda,\mu_B,A^0) > 0$$
$$\rightarrow p(A=a, B=b \mid \lambda,\mu_A,\mu_B) > 0.$$

Particle Compatibility:

$$\rho(\lambda \mid \phi,A,B) > 0 \leftrightarrow \rho(\lambda \mid \phi,A,B^0) > 0 \leftrightarrow \rho(\lambda \mid \phi,A^0,B^0) > 0.$$

Apparatus Compatibility:

(i) $\rho(\mu_A \mid \lambda,A,B^0) > 0$ and $\rho(\mu_B \mid \lambda,B,A^0) > 0$
$$\leftrightarrow \rho(\mu_A,\mu_B \mid \lambda,A,B) > 0.$$

(ii) If $\rho(\mu_A,\mu_B \mid \lambda,A,B)$ is finite, then so are $\rho(\mu_A \mid \lambda,A,B^0)$ and $\rho(\mu_B \mid \lambda,B,A^0)$.

Let's quickly compare these conditions to the corresponding Bell locality assumptions.

Particle Compatibility. Particle Compatibility *permits* nonlocal connections prohibited by Particle Locality, according to which $\rho(\lambda \mid \phi,A,B) = \rho(\lambda \mid \phi,A,B^0) = \rho(\lambda \mid \phi,A^0,B^0)$. Under the Particle Compatibility corollary, setting up an apparatus (or changing an apparatus setting) can make the particles more or less likely to occupy a

given state. The corollary demands only that setting up an apparatus, or changing its setting, not make it impossible for the particles to occupy a previously-possible λ. (In section 2.7.3, I'll explain why I equate "impossibility" with zero probability *density*.) So for instance, Particle Compatibility demands that if it's possible for the particles to occupy state λ when the B-apparatus is turned on, then it's also possible for the particles to occupy state λ when the B-apparatus is switched off.

Particle Locality trivially implies Particle Compatibility, but not vice versa.

Weak Factorizability and Apparatus Compatibility. These two conditions take roughly the following form: If some event on the A-wing of the EPR experiment has nonzero probability (density) when apparatus 2 is turned off, and some event on the B-wing has nonzero probability (density) when apparatus 1 is turned off, then those two events have nonzero probability (density) of happening together when both apparatuses are turned on. For instance, Weak Factorizability allows the A-measurement outcome to affect the probability of obtaining B=b on particle 2, in violation of regular Factorizability. In symbols, Weak Factorizability allows $p(B=b \mid A=a, \lambda, \mu_A, \mu_B) \neq p(B=b \mid \lambda, \mu_B, A^0)$. Weak Factorizability demands only that if $p(B=b \mid \lambda, \mu_B, A^0)>0$, then $p(B=b \mid A=a, \lambda, \mu_A, \mu_B)>0$. In words, given the fully-specified state of the particle pair, obtaining a measurement outcome on particle 1 cannot reduce to zero probability an otherwise-possible[11] measurement result for particle 2.

---

[11]Here, by "otherwise-possible," we mean a measurement result having nonzero-probability of occurring. Of course, a measurement outcome can be possible even when its probability is zero. In section 2.7.3, I'll treat these subtleties more carefully.

Similarly, Apparatus Compatibility allows the state of apparatus 2 to depend nonlocally on the state of apparatus 1. This violates TAF, according to which $\rho(\mu_A,\mu_B \mid \lambda,A,B) = \rho(\mu_A \mid \lambda,A,B^0)\cdot\rho(\mu_B \mid \lambda,A^0,B)$. Apparatus Compatibility requires only that the nonlocal connection (i) not be "strong" enough to render any (μ_A,μ_B) pair incompatible, and (ii) not be so strong that, by turning off apparatus 1, we can make a given apparatus 2 microstate *infinitely* more likely to occur than would have been the case had apparatus 1 remained on.

TAF implies Apparatus Compatibility, and Factorizability trivially implies Weak Factorizability. But both converses fail.

In summary, each of the three new locality conditions weakens the corresponding Bell locality assumption.

§2.7.2. Nonlocality theorem using the weakened conditions

I'll now prove

Theorem: Weak Factorizability & Particle Compatibility & Apparatus Compatibility → Contradiction with QM's predictions.

Proof: Recall the structure of my original nonlocality theorem in section 2.5.3 above. Invoking Particle Locality, I first proved (in appendix 2.5.5) that if the hidden-variable theory reproduces QM's predictions, then there exists a λ that

(i) mirrors all 192 QM perfect anticorrelations needed to complete the proof, and

(ii) mirrors the Spectrum Rule with respect to the 49 observables used in the proof.

I then showed that, in any stochastic Bell local theory, this λ generates an inconsistent map (by the Kochen-Specker-Peres contradiction).

My strategy here is similar. In the frightfully boring appendix at the end of this chapter, I'll show that if the hidden-variable theory reproduces QM's predictions, then Particle Compatibility implies the existence of a λ obeying (i) and (ii). For now, let me take this result as given. It only remains to show that, given such a λ, a Kochen-Specker contradiction follows from Weak Factorizability and Apparatus Compatibility.

To complete the proof, it will be useful for me to first prove a lemma. Specifically, I'll show that Weak Factorizability and Apparatus Compatibility imply a certain mathematical condition. From that condition, the Kochen-Specker contradiction will follow relatively quickly.

Lemma: If Weak Factorizability and Apparatus Compatibility hold, then the following implication holds:

$$p(A=a, B=b \mid \lambda)=0 \rightarrow p(A=a \mid \lambda,B^0)=0 \text{ or } p(B=b \mid \lambda,A^0)=0.$$

The proof of this lemma proceeds by contrapositive. Suppose that $p(A=a \mid \lambda,B^0)>0$ *and* $p(B=b \mid \lambda,A^0)>0$. That is, suppose

$$p(A=a \mid \lambda,B^0) \equiv \int p(A=a \mid \lambda,\mu_A,B^0)\cdot\rho(\mu_A \mid \lambda,A,B^0)\cdot d\mu_A > 0,$$

$$p(B=b \mid \lambda,A^0) \equiv \int p(B=b \mid \lambda,\mu_B,A^0)\cdot\rho(\mu_B \mid \lambda,B,A^0)\cdot d\mu_B > 0.$$

Therefore, there exists an "anomalous" set of μ_A-states, call it M_A, for which $\rho(\mu_A \mid \lambda,A,B^0)>0$ and $p(A=a \mid \lambda,\mu_A,B^0)>0$; and M_A is a nonzero-measure subset of $\{\mu_A\}$, the set of all microstates for which $\rho(\mu_A \mid \lambda,A,B^0)>0$. Similarly, there exists M_B, the set of μ_B-states for which $\rho(\mu_B \mid \lambda,B,A^0)>0$ and $p(B=b \mid \lambda,\mu_B,A^0)>0$; and M_B is a nonzero-measure subset of $\{\mu_B\}$.

Let $M_A x M_B$ denote the set of "anomalous" joint microstates formed by pairing each member of M_A with each member of M_B. Since M_A is a nonzero-measure subset of $\{\mu_A\}$ and M_B is a nonzero-measure subset of $\{\mu_B\}$, it follows that $M_A x M_B$ is a nonzero-measure subset of $\{\mu_A\}x\{\mu_B\}$ with respect to the measure $\rho(\mu_A \mid \lambda,A,B^0)\rho(\mu_B \mid \lambda,B,A^0)$. In symbols,

$$\int_{M_A x M_B} \rho(\mu_A \mid \lambda,A,B^0)\rho(\mu_B \mid \lambda,B,A^0)\cdot d\mu_A d\mu_B > 0.$$

Does it follow that $M_A x M_B$ also has nonzero measure with respect to $\rho(\mu_A,\mu_B \mid \lambda,A,B)$? Yes, and here's why. By Apparatus Compatibility part (i), if $\rho(\mu_A \mid \lambda,A,B^0)\rho(\mu_B \mid \lambda,B,A^0) > 0$, then $\rho(\mu_A,\mu_B \mid \lambda,A,B) > 0$. And by Apparatus Compatibility part (ii), if the product $\rho(\mu_A \mid \lambda,A,B^0)\rho(\mu_B \mid \lambda,B,A^0)$ "blows up" to infinity at some (μ_A, μ_B) pair, then so does $\rho(\mu_A,\mu_B \mid \lambda,A,B)$. More precisely, if $\rho(\mu_A \mid \lambda,A,B^0)\rho(\mu_B \mid \lambda,B,A^0)\cdot d\mu_A d\mu_B > 0$ (due to a "delta function,"), then $\rho(\mu_A,\mu_B \mid \lambda,A,B)d\mu_A d\mu_B > 0$. Roughly speaking, Apparatus Compatibility guarantees that for all (μ_A, μ_B) pairs, $\rho(\mu_A,\mu_B \mid \lambda,A,B)$ is always a finite fraction of $\rho(\mu_A \mid \lambda,A,B^0)\rho(\mu_B \mid \lambda,B,A^0)$. Therefore, by measure theory,

$$\int_{M_A x M_B} \rho(\mu_A,\mu_B \mid \lambda,A,B)\cdot d\mu_A d\mu_B > 0.$$

Since by construction $p(A=a \mid \lambda,\mu_A,B^0) > 0$ and $p(B=b \mid \lambda,\mu_B,A^0) > 0$ for all (μ_A, μ_B) pairs in $M_A x M_B$, it follows that

$$\int_{M_A x M_B} p(A=a \mid \lambda,\mu_A,B^0)\cdot p(B=b \mid \lambda,\mu_B,A^0)\cdot \rho(\mu_A,\mu_B \mid \lambda,A,B)\cdot d\mu_A d\mu_B > 0.$$

By Weak Factorizability, since $p(A=a \mid \lambda,\mu_A,B^0)\cdot p(B=b \mid \lambda,\mu_B,A^0) > 0$ for all (μ_A, μ_B) pairs in $M_A x M_B$, it follows that $p(A=a, B=b \mid \lambda,\mu_A,\mu_B) > 0$ for all (μ_A, μ_B) pairs in $M_A x M_B$. Since all these probabilities are *finite,* we don't have to worry about "blow ups," and hence it immediately follows that

$$\int_{M_A x M_B} p(A=a, B=b \mid \lambda,\mu_A,\mu_B)\cdot \rho(\mu_A,\mu_B \mid \lambda,A,B)\cdot d\mu_A d\mu_B > 0.$$

Since $M_A x M_B$ is a subset of all possible (μ_A,μ_B) pairs, we have

$$\int p(A=a, B=b \mid \lambda,\mu_A,\mu_B)\cdot \rho(\mu_A,\mu_B \mid \lambda,A,B)\cdot d\mu_A d\mu_B > 0.$$

But the left-hand side of this inequality is, by definition, $p(A=a, B=b \mid \lambda)$. So, we conclude that $p(A=a, B=b \mid \lambda) > 0$.

Assuming Weak Factorizability and Apparatus Compatibility, I just proved that if $p(A=a \mid \lambda,B^0)>0$ and $p(B=b \mid \lambda,A^0)>0$, then $p(A=a,B=b \mid \lambda)>0$. It follows that if

$p(A=a,B=b \mid \lambda)=0$, then $p(A=a \mid \lambda,B^0)=0$ or $p(B=b \mid \lambda,A^0)=0$. That's exactly the lemma I needed to prove. Q.E.D.

Given this lemma, and Kochen-Specker contradiction can easily be reached by reprising the reasoning of section 2.5.3. Remember, for now we're taking as given the result from the upcoming appendix that any theory reproducing QM's predications and obeying Particle Compatibility must contain a particle state λ that mirrors all of those perfect anticorrelations and also mirrors the spectrum rule with respect to the relevant spin-component and spin-Hamiltonian observables. Consider this λ.

Since λ obeys the spectrum rule with respect to a spin-Hamiltonian H_s, the only three H_s-measurement outcomes that can have nonzero probability are h_x, h_y, and h_z. Since probabilities are normalized,

$$p(H_s=h_x \mid \lambda,S^0) + p(H_s=h_y \mid \lambda,S^0) + p(H_s=h_z \mid \lambda,S^0) = 1.$$

("S^0" denotes that no measurement occurs on particle 2.) It follows that at least one of those three spin-Hamiltonian measurement-result probabilities is greater than zero.

- Suppose $p(H_s=h_x \mid \lambda,S^0) > 0$.

From QM, we have the following four perfect anticorrelations:

(a) $P_{QM}(H_s=h_x, S_x=+1 \mid \Psi_{singlet}) = 0,$

(a') $P_{QM}(H_s=h_x, S_x=-1 \mid \Psi_{singlet}) = 0,$

(b) $P_{QM}(H_s=h_x, S_y=0 \mid \Psi_{singlet}) = 0,$

(c) $P_{QM}(H_s=h_x, S_z=0 \mid \Psi_{singlet}) = 0.$

Since the λ under consideration reproduces these anticorrelations,

(a) $p(H_s=h_x, S_x=+1 \mid \lambda) = 0,$

(a') $p(H_s=h_x, S_x=-1 \mid \lambda) = 0,$

(b) $p(H_s=h_x, S_y=0 \mid \lambda) = 0,$

(c) $p(H_s=h_x, S_z=0 \mid \lambda) = 0.$

From the lemma just proven, we immediately get

(a) $p(H_s=h_x \mid \lambda,S^0) = 0$ or $p(S_x=+1 \mid \lambda,H^0) = 0$

(a') $p(H_s=h_x \mid \lambda,S^0) = 0$ or $p(S_x=-1 \mid \lambda,H^0) = 0$

(b) $p(H_s=h_x \mid \lambda,S^0) = 0$ or $p(S_y=0 \mid \lambda,H^0) = 0,$

(c) $p(H_s=h_x \mid \lambda,S^0) = 0$ or $p(S_z=0 \mid \lambda,H^0) = 0.$

(H^0 denotes that no measurement occurs on particle 1.) By supposition, $p(H_s=h_x \mid \lambda,S^0) > 0$. Therefore

(a) $p(S_x=\pm 1 \mid \lambda,H^0) = 0$

(b) $p(S_y=0 \mid \lambda,H^0) = 0,$

(c) $p(S_z=0 \mid \lambda,H^0) = 0.$

Since λ obeys the spectrum rule with respect to S_x, normalization implies

$$p(S_x=-1 \mid \lambda,H^0) + p(S_x=0 \mid \lambda,H^0) + p(S_x=+1 \mid \lambda,H^0) = 1.$$

From this and (a), we immediately get $p(S_x=0 \mid \lambda,H^0) = 1$. In summary, we have

(a) $p(S_x=0 \mid \lambda,H^0) = 1$

(b) $p(S_y=0 \mid \lambda,H^0) = 0,$

(c) $p(S_z=0 \mid \lambda,H^0) = 0.$

This conclusion, for the particles in state λ, followed from the above lemma (proven by assuming Weak Factorizability and Apparatus Compatibility), and from the supposition that $p(H_s=h_x \mid \lambda,S^0) > 0$. If we suppose instead that $p(H_s=h_y \mid \lambda,S^0) > 0$, reasoning similar to the above, with **x**, **y**, and **z** cyclically permuted, yields $p(S_x=0 \mid \lambda,H^0)=0$, $p(S_y=0 \mid \lambda,H^0)=1$, and $p(S_z=0 \mid \lambda,H^0)=0$. Similarly, if we suppose $p(H_s=h_z \mid \lambda,S^0) > 0$, we conclude that $p(S_x=0 \mid \lambda,H^0)=0$, $p(S_y=0 \mid \lambda,H^0)=0$, and $p(S_z=0 \mid \lambda,H^0)=1$.

As noted above, by the spectrum rule, at least one of those three spin-Hamiltonian measurement-result probabilities is greater than 0. Therefore, from the previous paragraph, we see that the three values

$$\{p(S_x=0 \mid \lambda,H^0),\ p(S_y=0 \mid \lambda,H^0),\ p(S_z=0 \mid \lambda,H^0)\}$$

must be such that two of the values equal 0 while the third value equals 1.

Due to the spherical symmetry of the spin singlet state Ψ, the same conclusion applies to *each* of the 16 orthogonal triads of directions needed to generate the Kochen-Specker-Peres contradiction. Now all we have to do is map the value $p(S_n=0 \mid \lambda,H^0)$ to the point **n** on the unit sphere, for all **n** used in this proof. As discussed in sections 2.4 and 2.5, this map is algebraically impossible, by the Kochen-Specker-Peres contradiction. This contradiction establishes that no theory obeying the weakened locality conditions (Particle Compatibility, Apparatus Compatibility, and Weak Factorizability) can reproduce the perfect anticorrelations of QM. *Q.E.D.*

In summary: As shown in the upcoming appendix at the end of this chapter, Particle Compatibility implies the existence of a λ that reproduces the relevant perfect anticorrelations and mirrors the spectrum rule for the relevant observables, assuming the hidden-variable theory reproduces QM's predictions. By the Kochen-Specker contradiction, such a λ is inconsistent with the mathematical condition introduced in the above lemma: $p(A=a, B=b \mid \lambda)=0 \rightarrow p(A=a \mid \lambda,B^0)=0$ or $p(B=b \mid \lambda,A^0)=0$. This condition follows from Weak Factorizability and Apparatus Compatibility. Putting all this together, we see that no theory consistent with QM can obey my three weakened locality assumptions (Particle Compatibility, Apparatus Compatibility, and Weak Factorizability).

§2.7.3. Philosophical implications: Zero probability vs. impossibility

In the following sections, I'll explore the philosophical ramifications of theorem 2.7.2. To do so, I must first review the connections between zero-probability and

impossibility. Then, I'll show that the three weakened locality conditions used in theorem 2.7.2 are motivated by weakened Bell locality, a metaphysical constraint less stringent than regular Bell locality. Because the theorem suggests that nature violates weakened Bell locality, I'll explore the physical and metaphysical content of this constraint.

An event can be possible even though its probability of occurring is zero. To see why, imagine choosing a random real number between zero and one. It's possible that you'll pick 0.6. Indeed, that number is as likely as any other. Put more technically, the probability *density* of getting 0.6 equals the probability density of getting any other number between zero and one: $\rho(0.6)=1$. Nonetheless, the *probability* of choosing 0.6 is zero, because 0.6 is one of an uncountably infinite number of possible results.

By contrast, in this game, choosing the number -0.6 is impossible. Mathematically, this corresponds to the fact that getting -0.6 not only has zero probability, but also has zero probability *density*: $\rho(-.6)=0$.

This game illustrates two related points. First, in "standard" cases, a possible event *a* has zero probability *because* it is one of an uncountably infinite number of possible events.

Second, an event is impossible if and only if it has zero probability *density* of occurring. For if an event has nonzero probability density, then it would have nonzero probability of occurring were the relevant "game" repeated an uncountably infinite number of times; hence the event is possible. And if an event has zero probability density, then it would occur zero times, even if the game were repeated an infinite number of times. As P. Suppes (personal conversation with Martin Jones) notes, most

probability theorists agree that it's unproblematic to associate zero probability density with impossibility

Now consider a game in which only a finite number (or at most a countable infinity) of results are possible. As a simple example, imagine a machine that prints out one of two numbers, either 100 or 101. No other result is possible. And furthermore, suppose that obtaining the result 100 has zero probability. Does it follow that getting 100 is impossible? The answer depends on the inner workings of the machine. If the laws describing those inner working are such that 100 simply cannot be obtained, then 100 is indeed impossible. But suppose the machine, as an intermediate stage, picks a random real number between zero and one; and then the machine prints out "100" if that random number is 0.6, and prints out "101" otherwise. In this weird case, obtaining 100 is indeed possible, even though that result has zero probability.

My point is this: When the relevant "game" has a finite number of possible results, one of those results can have zero probability, but only in specially-contrived cases. In standard cases, we expect that if a is one of a finite number of possible results, then a will have nonzero probability.

§2.7.4. Philosophical implications: Weakened Bell locality

In this subsection, I introduce weakened Bell locality, and show that this requirement motivates Weak Factorizability, Particle Compatibility, and Apparatus Compatibility, the three weakened locality conditions of theorem 2.7.2. I reproduce these conditions for easy reference:

Weak Factorizability:

$p(A=a \mid \lambda,\mu_A,B^0) > 0$ and $p(B=b \mid \lambda,\mu_B,A^0) > 0$

$$\rightarrow p(A=a, B=b \mid \lambda,\mu_A,\mu_B) > 0.$$

Particle Compatibility:

$$\rho(\lambda \mid \phi,A,B) > 0 \leftrightarrow \rho(\lambda \mid \phi,A,B^0) > 0 \leftrightarrow \rho(\lambda \mid \phi,A^0,B^0) > 0.$$

Apparatus Compatibility:

(i) $\rho(\mu_A \mid \lambda,A,B^0) > 0$ and $\rho(\mu_B \mid \lambda,B,A^0) > 0$

$$\leftrightarrow \rho(\mu_A,\mu_B \mid \lambda,A,B) > 0.$$

(ii) If $\rho(\mu_A,\mu_B \mid \lambda,A,B)$ is finite, then so are $\rho(\mu_A \mid \lambda,A,B^0)$ and $\rho(\mu_B \mid \lambda,B,A^0)$.

Suppose *a* and *b* are spacelike separated events. Then we have

WEAKENED BELL LOCALITY: An event *a* cannot affect the possibility of a spacelike separated event. Specifically, if *b* is possible when *a* does not occur, then *b* is possible when *a* does occur. Also, if *b* is impossible when *a* does not occur, then *b* is impossible when *a* does occur.

Put another way, *a* cannot render impossible a spacelike separated event that otherwise might have occurred. Nor can *a* render possible a spacelike separated event that otherwise could not have occurred.

Because events are possible just in case they have nonzero probability density, weakened Bell locality logically implies both Particle Compatibility and Apparatus Compatibility part (i).[12]; if a particle or apparatus state has nonzero probability density of obtaining when an apparatus is off (on), then a spacelike separated event--the switching off or on of a distant apparatus--cannot "rule out" that state.

Weakened Bell locality, however, does not entail Weak Factorizability or Apparatus Compatibility part (ii) . Weak Factorizability demands that obtaining an A-measurement outcome not reduce to zero the probability of getting a certain B-measurement result. As discussed above, a zero-probability event can be possible. Consequently, a theory violates Weak Factorizability *without* violating weakened Bell locality if, and only if, the theory asserts that a joint measurement result A=a & B=b is possible even though it has zero probability.

Similar considerations apply to Apparatus Compatibility part (ii). If that condition fails, then an apparatus microstate with infinite probability density (and hence, nonzero *probability* of occurring) can have its probability reduced to zero by the switching on of a distant apparatus. Since that zero-probability microstate still has nonzero probability *density* and is therefore still *possible,* a theory could incorporate this feature without violating weakened Bell Locality.

In the remainder of this subsection, I argue that such a theory is highly contrived and physically implausible.

Let's start with Weak Factorizability. The A's and B's used in the above nonlocality theorems are spin-component and spin-Hamiltonian observables on spin-1 particles.

---

[12]I'm assuming no "conspiracies."

These observables are discrete. According to any theory consistent with QM, the probability is zero that measurement will yield a non-eigenvalue; and these spin observables have a finite number of eigenvalues. Therefore, if a zero-probability joint measurement result A=a & B=b is possible, it's *not* because A=a & B=b is one of an infinite number of possible results. This experiment is *not* analogous to picking a random real number between zero and one. Rather, in order to incorporate a zero-probability yet possible result A=a & B=b, a theory must rely on some contrivance, analogous to the "100" machine described above.

For instance, consider particle and apparatus states such that $p(A=a,B=b \mid \lambda,\mu_A,\mu_B)=0$. A hidden-variable theorist could claim the following: When A and B undergo measurement, nature picks out a random real number between zero and one. If that number is .6, then the measuring devices record "A=a" and "B=b." Otherwise, the measuring devices record another pair of outcomes. In this theory, the measurement result A=a & B=b is possible, even though it has zero probability. This theory, however, is artificial and implausible.

Alternatively, the hidden-variable theory could simply declare, as a first principle, that A=a & B=b is possible even though its probability is zero. This move seems *ad hoc*.

In less contrived theories, a perfect anticorrelation usually reflects an underlying conservation law, as discussed in section 2.6 above. For instance, in QM, the perfect anticorrelations invoked in theorem 2.7.2 follow ultimately from rotational symmetry, which leads to conservation of angular momentum. Conservation of angular momentum is considered to be a fundamental law. Therefore, in QM, we can assert

with full counterfactual force that a (non-erroneous) joint measurement result directly contradicting conservation of angular momentum *would not* occur. In other words, for the A's and B's invoked in theorem 2.7.2, correctly obtaining A=a & B=b when $p(A=a, B=b \mid \Psi_{singlet})=0$ is impossible, according to QM. This conclusion applies to all theories in which the relevant perfect anticorrelations follow from fundamental conservation laws.

In Bohm's theory, too, perfect anticorrelations are impossible to violate. Bohm's theory is deterministic. Under determinism, events evolve inexorably. All future states of affairs, except the one pre-determined by the initial conditions, *could* not occur. Therefore, when the complete state of the universe is such that $p(A=a,B=b \mid \lambda,\mu_A,\mu_B)=0$, obtaining A=a & B=b is physically impossible. This conclusion applies to all deterministic theories.

So, far, I've shown that only a contrived theory would violate Weak Factorizability without also violating weakened Bell locality. A similar though less "clean" argument applies to Apparatus Compatibility part (ii). Suppose that condition fails. Then there exists some μ_A's such that $\rho(\mu_A \mid \lambda,A,...)$ is infinite when apparatus 2 is turned off, but finite when apparatus 2 is turned on. So, there must be an uncountable infinity of μ_A's. An infinite $\rho(\mu_A \mid \lambda,A,...)$ means that μ_A has nonzero probability of occurring. But as just discussed, usually when an event has nonzero probability, it's because the event is one of a countable number that could occur. To violate Apparatus Compatibility part (ii) without violating weakened Bell locality, a theory has to do more than simply introduce a state space $\{\mu_A\}$ in which a bunch of discrete delta-function "spikes" stick up out of the background "soup" of finite-probability density μ_A's. The theory must also incorporate a

nonlocal interaction strong enough--and contrived enough--to "shrink" those spikes back into the soup when apparatus 2 gets turned on. Furthermore, in shrinking the "spiked" μ_A probability densities by a factor of infinity, the interaction may not also shrink by an infinite factor the "non-spiked" μ_A probability densities, because doing so would reduce those probability densities to 0, in violation of Apparatus Compatibility part (i). So, this ρ-shrinking interaction would have to be miraculously selective. Indeed, it would be hard for interaction terms to have these properties unless they were specifically constructed with that purpose. For this reason, only a contrived theory would violate Apparatus Compatibility part (ii) without also violating weakened Bell locality.

Since the argument of this subsection is messy, let me summarize it. Weakened Bell locality entails Particle Compatibility and Apparatus Compatibility part (i). Therefore, *any* theory violating either of those conditions automatically violates weakened Bell locality. By contrast, a theory could conceivably violate Weak Factorizability or Apparatus Compatibility part (ii) without violating weakened Bell locality. Such a theory must claim, for instance, that some of its perfect anticorrelations--i.e., some of its zero-probability joint measurement results--are possible. The perfect anticorrelations considered here involve observables for which only a finite number of outcomes have nonzero probability. (These experiments do not resemble choosing a random real number.) Therefore, some contrivance would be needed to ensure that zero-probability joint measurement results could occur. For instance, the theory could claim that the final measurement outcome depends on an "intermediate result," where the intermediate result corresponding to the perfect anticorrelation is one of an infinity of possible intermediate results. In my view, such contrivances, and also the contrivances

needed to "escape" Apparatus Compatibility without violating weakened Bell locality, seem artificial and physically implausible. In brief, only a highly contrived theory would violate my weakened locality assumptions without violating Weakened Bell locality. In this sense, Weakened Bell locality is the "guiding principle" behind Weak Factorizability, Particle Compatibility, and Apparatus Compatibility.

§2.7.5. Bell locality vs. weakened Bell locality

Throughout this subsection, I'll assume that no "contrivances" of the kind discussed above actually obtain. In this case, assuming QM's predictions hold, theorem 2.7.2 implies that nature violates weakened Bell locality. We now briefly explore the philosophical implications of this result.

Bell inequalities and previous algebraic nonlocality theorems suggest only that nature violates (stochastic) Bell locality. Bell locality requires that an event a not affect the *probability* of spacelike separated event b. By contrast, weakened Bell locality makes the less stringent demand that event a not affect the *possibility* of spacelike separated event b. To explore the physical difference between regular and weakened Bell locality, suppose that Bell locality fails while weakened Bell locality holds. Then we can picture the world as follows: Events evolve in spacetime, constrained by certain rules. The set of events $\{a\}$ that could possibly occur in spacetime region **R** is determined entirely by events that occurred in the backwards light cone of **R**.[13] The

---

[13]We mean "event" in its broadest sense. For instance, we call the state of all the objects in region **R**--or if you prefer, the state of **R**--an event.

probability (density) however, that a given element of $\{a\}$ occurs depends also on events spacelike separated from **R**. So, a nonlocal connection can "tweak" the likelihood that a specific element of $\{a\}$ occurs. But no nonlocal connection is "strong" enough to alter $\{a\}$.

By contrast, if weakened Bell locality fails, then $\{a\}$ itself is determined, in part, by spacelike separated events.

With this said, I'll now admit that this distinction between regular and weakened Bell locality carries limited metaphysical significance. Local connections between events are capable of "ruling out" some events. If Bell-nonlocal connections exist, why should they be any less capable of ruling out events? In other words, if Bell locality fails, why shouldn't weakened Bell locality also fail?

Also, notice that if determinism holds, then regular and weakened Bell locality are equivalent. Under determinism, changing an event's probability (say, from zero to one) is tantamount to changing its possibility (in this case, from impossible to possible). Therefore, with respect to deterministic theories, theorem 2.7.2 does not force us to accept *new* philosophical consequences.

Nonetheless, I think theorem 2.7.2 is worthwhile, not just because the physical distinction between regular and weakened Bell locality is kind of interesting (for stochastic theories), but also because some philosophers may try to attach more metaphysical significance to this distinction. Also, this result shows for the first time that algebraic (perfect correlations) nonlocality proofs are *better* in a sense than regular Bell-type statistical arguments, in the sense that algebraic proofs can get by with weaker assumptions.

Summary: Because zero-probability events can be possible, weakened Bell locality does not imply all three of our weakened locality conditions. But in order to violate Weak Factorizability without also violating weakened Bell locality, a theory would have to incorporate a physically-implausible contrivance. For this reason, weakened Bell locality strongly motivates our weakened locality assumptions. Since these assumptions imply a contradiction with QM's predictions, we have strong reason to think that nature violates weakened Bell locality. This violation forces us to accept that spacelike separated events not only affect each other's probabilities, but also affect each other's possibilities.

§2.7.6. APPENDIX: Part of theorem 2.7.2

If a hidden-variable theory reproduces QM's predictions, then Particle Compatibility implies the existence of a λ that
(i) mirrors all 192 QM perfect anticorrelations used in theorem 2.7.2; and
(ii) mirrors the spectrum rule with respect to the 49 observables used in that theorem.

Proof: Let $P_{QM}(A_i=a_i, B_i=b_i \mid \Psi)=0$ denote any one of the relevant perfect anticorrelations or spectrum rule occurrences used in theorem 2.7.2. For instance, $A_1=H_s$, $a_1=h_x$, $B_1=S_x$, and $b_1=+1$; $A_2=H_s$, $a_2=h_x$, $B_2=S_y$, and $b_2=0$; and so on. To mirror the Spectrum Rule with respect to H_s, set $A_{193}=H_s$, $a_{193}=\{$non-eigenvalues of $H_s\}$, and $B_{193}=S^0$, where S^0 indicates that no measurement occurs on particle 2. Crucially, we'll need to consider only a finite number of i's.

Since the hidden variable theory reproduces those perfect anticorrelations and spectrum rule occurrences,

$$(*) \qquad 0 = p(A_i=a_i,\, B_i=b_i \mid \Psi) = \int_{\{\lambda\}} p(A_i=a_i,\, B_i=b_i \mid \lambda)\cdot\rho(\lambda \mid \Psi, A_i, B_i)\cdot d\lambda,$$

for all i, where $\{\lambda\}$ is the set of λ states for which $\rho(\lambda \mid \Psi,A_i,B_i)>0$. By Particle Compatibility, $\rho(\lambda \mid \Psi,A_i,B_i)>0$ iff $\rho(\lambda \mid \Psi,A^0,B^0)$, and hence $\{\lambda\}$ is the same for all i. Remember, $\rho(\lambda \mid \Psi,A^0,B^0)$ denotes the probability density when no measurement occurs on either particle.

To complete this proof, I'll consider two cases: (1) $\{\lambda\}$ contains a finite or countably infinite number of members, and (2) $\{\lambda\}$ contains an uncountable infinity of members.

Case 1: $\{\lambda\}$ finite or countably infinite. Then eq. (*) becomes a finite or infinite sum, and the nonzero probability density becomes a nonzero probability:

$$(**) \qquad 0 = p(A_i=a_i,\, B_i=b_i \mid \Psi) = \sum_\lambda p(A_i=a_i,\, B_i=b_i \mid \lambda)\cdot p(\lambda \mid \Psi, A_i, B_i),$$

where the sum is only over λ states for which $p(\lambda \mid \Psi,A_i,B_i)>0$. If we assume that a probability measure on a countable number of elements never assigns nonzero probability density except when it assigns nonzero probability, then the sum is over the states in $\{\lambda\}$. As discussed in section 2.7.3, this extra assumption is valid. When the number of elements is uncountably infinite, then an element can have nonzero probability density but zero probability. But here, it's vacuous to say that $\rho(\lambda \mid$

$\Psi,A_i,B_i)>0$ even though $p(\lambda \mid \Psi,A_i,B_i)>0$, because if $p(\lambda \mid \Psi,A_i,B_i)>0$, then that λ state doesn't contribute to sum. Therefore, even though $\rho(\lambda \mid \Psi,A_i,B_i)>0$, that λ state in no sense "contributes" to the hidden-variable states underlying the quantum state.

From eq. (**), it follows that for all λ states in $\{\lambda\}$, $p(A_i=a_i, B_i=b_i \mid \lambda)=0$. As noted above, by Particle Compatibility, $\{\lambda\}$ is the same for all i. So, for all i and for all members of $\{\lambda\}$, $p(A_i=a_i, B_i=b_i \mid \lambda)=0$. In words, the elements of $\{\lambda\}$ reproduce all the perfect anticorrelations and spectrum rule occurrences needed to complete theorem 2.7.2. This completes the argument for the finite or countably infinite case.

Case 2: $\{\lambda\}$ uncountably infinite. By Particle Compatibility, $\rho(\lambda \mid \Psi,A^0,B^0)\geq 0$ iff $\rho(\lambda \mid \Psi,A_i,B_i)>0$; and $\{\lambda\}$ denotes the set of states for which $\rho(\lambda \mid \Psi,A^0,B^0)>0$.

From eq. (*) above, $p(A_i=a_i, B_i=b_i \mid \lambda)=0$ almost everywhere in $\{\lambda\}$. It follows from measure theory (cf. Wheedan and Zygmund 1977) that for any finite probability density ρ, $\int_{\{\lambda\}} p(A_i=a_i, B_i=b_i \mid \lambda)\cdot\rho\cdot d\lambda = 0$. I want to reach the conclusion that $\int_{\{\lambda\}} p(A_i=a_i, B_i=b_i \mid \lambda)\cdot\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda = 0$. But this doesn't immediately follow, because $\rho(\lambda \mid \Psi,A^0,B^0)$ might not be finite. That is, there might exist λ's such that $\rho(\lambda \mid \Psi,A^0,B^0)$ is infinite, i.e., $\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda > 0$. So, my strategy is to show that, "at worst," $\rho(\lambda \mid \Psi,A^0,B^0)$ blows up at a countable number of λ's. Hence, we can "subtract off" from $\{\lambda\}$ the states for which $\rho(\lambda \mid \Psi,A^0,B^0)$ blows up, and we'll be left with a nonzero-measure set of states with respect to all relevant measures. It will then be easy to complete the argument.

Let Λ_{blow} denote the subset of $\{\lambda\}$ whose probability densities blow up. Formally, Λ_{blow} contains the states λ such that $\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda > 0$. So, the states in Λ_{blow} have

nonzero probability of occurring, and that probability is $p(\lambda \mid \Psi,A^0,B^0) = \rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda$.

It should be intuitively clear that Λ_{blow} contains only a countable number of members. Since $\int_{\{\lambda\}}\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda = 1$, it follows that $\int_{\Lambda_{blow}}\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda \leq 1$. But $\int_{\Lambda_{blow}}\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda$ is really a sum over the elements of Λ_{blow}:

$$\int_{\Lambda_{blow}}\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda = \sum_{\Lambda_{blow}} p(\lambda \mid \Psi,A^0,B^0),$$

where all the $p(\lambda \mid \Psi,A^0,B^0)$ are greater than 0. From measure theory, a sum of positive nonzero numbers is finite only if the number of terms in the sum is at most countably infinite. In other words, the sum $\sum_{\Lambda_{blow}} p(\lambda \mid \Psi,A^0,B^0)$ would blow up if Λ_{blow} contained an uncountably infinite number of members.

Since Λ_{blow} contains at most a countably infinite number of members, $\int_{\Lambda_{blow}} d\lambda = 0$. Therefore, since $\int_{\{\lambda\}} d\lambda > 0$, the "remainder" set $\{\lambda\} - \Lambda_{blow}$ obeys $\int_{\{\lambda\}-\Lambda_{blow}} d\lambda > 0$. (This reasoning assumes that the hidden-variable states have a "volume measure" given by $d\lambda$, and not just a ρ-measure given by $\rho(\lambda \mid \Psi,A^0,B^0)\cdot d\lambda$. A hidden-variable theory whose uncountably infinite states "live" in a strange space that allows no volume-measure could escape our proofs.)

Since $\int_{\{\lambda\}-\Lambda_{blow}} d\lambda > 0$, it follows that $\int_{\{\lambda\}-\Lambda_{blow}}\rho(\lambda \mid \Psi,A_i,B_j) > 0$ for all i. Here's the proof, suggested by Tim Callahan. For integer n, let Λ_n denote the subset of $\{\lambda\}-\Lambda_{blow}$ all of whose members satisfy $\rho(\lambda \mid \Psi,A_i,B_j) > 1/n$. Could it be the case that for all n, $\int_{\Lambda_n} d\lambda = 0$? In other words, could all the Λ_n have zero volume measure? If so, then $U_{n=1 \text{ to } \infty}\Lambda_n$, the union of the Λ_n sets for all n, also has zero volume measure, since the union

of a countable number of zero-measure sets is itself a zero-measure set. But $U_{n=1 \text{ to } \infty}\Lambda_n$ *is* $\{\lambda\}$–Λ_{blow}, which does not have zero volume measure. This contradiction establishes that for some n, Λ_n has nonzero volume measure with respect to $\rho(\lambda \mid \Psi,A_i,B_j)$. In symbols, for some n, $\int_{\Lambda_n} d\lambda = \varepsilon$ for some $\varepsilon > 0$. And by construction, $\rho(\lambda \mid \Psi,A_i,B_j) >$ 1/n for each element of Λ_n. Therefore, $\int_{\Lambda_n} \rho(\lambda \mid \Psi,A_i,B_j) \cdot d\lambda > \varepsilon/n$. In words, for some n, Λ_n has nonzero measure with respect to $\rho(\lambda \mid \Psi,A_i,B_j)$. Since Λ_n is a subset of $\{\lambda\}$–Λ_{blow}, it follows that $\{\lambda\}$–Λ_{blow} has nonzero measure with respect to $\rho(\lambda \mid \Psi,A_i,B_j)$. And this is true for all i. By the exact same reasoning, it's also true that $\{\lambda\}$-Λ_{blow} has nonzero measure with respect to $\rho(\lambda \mid \Psi,A^0,B^0)$. Just run the argument of this paragraph, everywhere substituting "$\rho(\lambda \mid \Psi,A^0,B^0)$" for "$\rho(\lambda \mid \Psi,A_i,B_j)$."

Now we're home free. From eq. (*) above we get

$$\int_{\{\lambda\}-\Lambda_{blow}} p(A_i=a_i, B_i=b_i \mid \lambda) \cdot \rho(\lambda \mid \Psi,A_i,B_j) \cdot d\lambda = 0.$$

Since $\{\lambda\}$–Λ_{blow} has nonzero measure with respect to $\rho(\lambda \mid \Psi,A_i,B_j)$, it follows that $p(A_i=a_i, B_i=b_i \mid \lambda) = 0$ almost everywhere in $\{\lambda\}$–Λ_{blow}. Since $\rho(\lambda \mid \Psi,A^0,B^0)$ is a finite measure on $\{\lambda\}$–Λ_{blow} -- that was the whole point of "subtracting off" Λ_{blow} -- it immediately follows from measure theory that

$$\int_{\{\lambda\}-\Lambda_{blow}} p(A_i=a_i, B_i=b_i \mid \lambda) \cdot \rho(\lambda \mid \Psi,A^0,B^0) \cdot d\lambda = 0,$$

for all i. Since (as shown above) $\{\lambda\}$–Λ_{blow} has nonzero measure with respect to $\rho(\lambda \mid \Psi,A^0,B^0)$, it follows that for each i, a measure-1 subset of $\{\lambda\}$–Λ_{blow} with respect to $\rho(\lambda$

$|\Psi,A^0,B^0)$ obeys $p(A_i=a_i, B_i=b_i \,|\, \lambda) = 0$. In words, for any given i, a measure-1 subset of the λ states in $\{\lambda\}-\Lambda_{blow}$ obey the corresponding perfect anticorrelation or spectrum rule occurrence. Since the algebraic nonlocality proof under consideration uses a finite number of perfect anticorrelations and spectrum rule occurrences (i.e., we're considering a finite number of i's), we can take the intersection of those measure-1 subsets, and the result is itself a measure-1 subset of $\{\lambda\}-\Lambda_{blow}$. All the λ states in that "intersection set" obey *all* the relevant perfect anticorrelations and spectrum rule occurrences. *Q.E.D.*

CHAPTER 3: NON-INVASIVE MEASURABILITY AND SQUIDs

Section 3.1: Introduction

In chapter 2, I added my contributions to the argument that nature disobeys local causality. But from those arguments, we can't tell whether nature violates local causality *because* of "causal" action at a distance, superluminal causal mediation, or a holistic, nonseparable connection between "different" objects. I'll press harder on this causality vs. holism distinction in chapter 5. Here, I'll use SQUIDs to argue that holism is the culprit. My conclusion will emerge from an extended discussion concerning the following questions: Do macroscopic systems, like the microscopic systems considered in chapter 2, disobey some kind of local causality? And if so, what does it tell us about nature?

Motivating this discussion is the observation that local causality no-go theorems apply to microscopic systems such as electron pairs and photon pairs. Perhaps all *macroscopic* systems can be described by a more "classical" theory. If so, then under certain metaphysical assumptions, macroscopic reality isn't infected by quantum weirdness.[14] Indeed, it's well known that "decoherent" interactions between a

[14]If one takes the "naive realistic" position that any such theory of macroscopic reality must ultimately reduce to the "fundamental" theory of microscopic reality, then a "classical" theory of macroscopic reality wouldn't be as metaphysically exciting, because the underlying fundamental theory of all reality--including macroscopic reality--would incorporate violation of local causality. Although macroscopic reality would hide those violations, they'd still be lurking beneath the surface.

By contrast, some antirealists believe that theories don't really "get at" reality, but instead give us a bastardized, veiled version of what's out there, a version filtered through our experimental, theoretical, and perhaps cultural biases. Within this framework, two theories describing different domains need not reduce to one another or to a "fundamental" theory that subsumes them both. So for instance, a theory of macroscopic

macroscopic system and its environment cause the system's density operator to quickly "reduce" to the classically-expected mixture. I'll discuss the interpretation of such results *ad nauseam* in chapter 4. For now, I'll focus on macroscopically "coherent" systems such as Superconducting Quantum Interference Devices (SQUIDs) and superfluids, systems that keep their weird "interference" properties for an appreciable time before "succumbing" to environmentally-induced decoherence. Can such systems also be described by a "classical" (hidden-variable) theory devoid of nonlocal/holistic connections and other examples of quantum weirdness?

As Leggett (1986a,b) shows by considering hypothetical SQUID experiments, the answer is "no." Leggett and Garg (1985) derive a "temporal Bell inequality" that's violated by any theory consistent with QM's statistical predictions, an inequality than can perhaps be tested in the lab (see Tesche 1990). But Leggett and I disagree about *why* macroscopic reality violates this inequality (assuming QM's predictions hold). According to Leggett, it's *because* SQUIDs violate, "Macrorealism," which requires all macroscopic objects to *possesses* certain macroscopic properties at all times. In this chapter, I'll argue that Leggett's conclusion is unwarranted. I'm not claiming that Macrorealism holds. I'm claiming only that SQUID experiments have little to tell us about Macrorealism. But Leggett-style SQUID experiments *can* rule out "Non-invasive measurability," the requirement that it be possible, at least in principle, to measure an object without disturbing its state more than a tiny bit. After establishing this result, I'll argue that violation of non-invasive measurability indicates the existence of holistic connections between "different" objects.

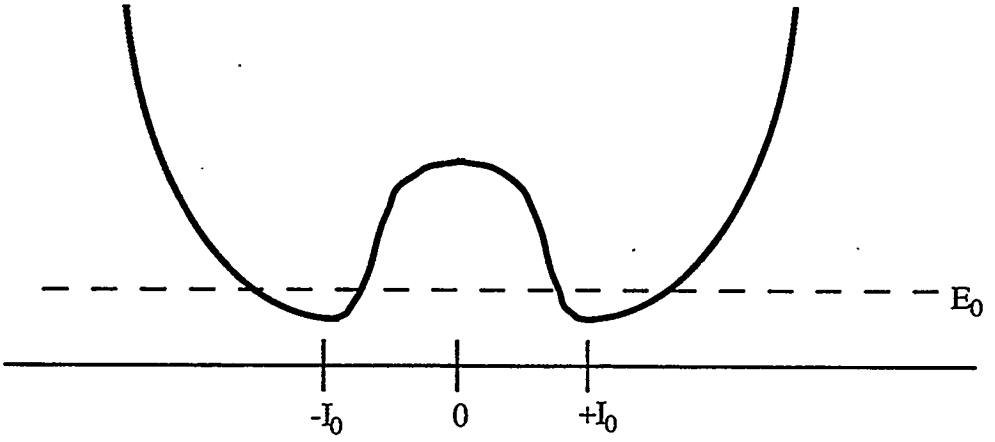
---

reality need not reduce to a theory of microscopic reality. The two theories could be on "equal footing," with neither more fundamental than the other, because neither theory is taken to be a fundamental description of reality *itself*. In this framework, the microscopic and macroscopic regimes can in principle be considered separately, and hence, a "classical" theory of all macroscopic reality would be dramatic.

Here's the game plan. In section 3.2, I'll lay out some formal details about SQUIDs. Then, in section 3.3, I'll present Leggett's derivation of a temporal Bell inequality, and I'll discuss the philosophical implications. In particular, I'll poke holes in his argument that violation of the inequality implies failure of Macrorealism. Finally, in section 3.4, I'll present my own derivation of Leggett's temporal Bell inequality, a derivation that relies on conditions significantly weaker than those used by Leggett. Specifically, my derivation does *not* assume Macrorealism. Using the new technical result, I'll argue that violation of the temporal Bell inequality has nothing to say about Macrorealism, but strongly suggests that nature violates non-invasive measurability.

Section 3.2: SQUID formalities

An rf SQUID consists of a superconducting ring (often several millimeters in diameter) containing a single Josephson junction. According to quantum mechanics, the Cooper-paired electrons are all in the same state, forming a Bose "gas." So, we can treat the SQUID's current as a single macroscopic parameter, call it *I*. For the SQUIDs under consideration, the currents are typically on the order of milliamps, which is indeed macroscopic both in terms of the number of electrons involved and in terms of easy detectability. In the absence of the Josephson junction, the SQUID's current eigenstates would correspond to integral Planck-units of magnetic flux threading through the loop, where the magnetic field is created by the current itself. This is still true when we insert the Josephson junction, which, in very rough terms, inserts an energy barrier between clockwise-current states and counterclockwise-current states. Leggett (1986b) works through the details. For my purposes (and Leggett's purposes), the technical minutiae aren't important. What's important is that, when the rf SQUID is placed in a properly-tuned external magnetic field, the effective potential as a function of current looks roughly like this:



Qualitatively, the energy eigenstates are similar to those found in a box-with-a-barrier-in-the-middle potential or an ammonia molecule (inversion state), but with a major difference: The SQUID's current, unlike a boxed particle's position, is quantized, since the magnetic flux through the SQUID ring is quantized. I_0 denotes the "quantum" of current. By tweaking the relevant parameters (size of ring, properties of Josephson junction), we can ensure that $I=\pm I_0$ lies near the bottom of the left and right energy wells, as drawn above.

Let $|+\rangle$ and $|-\rangle$ denote the $I=+I_0$ and $I=-I_0$ eigenstates of the current operator, which I'll call **Q** so as not confuse it with the identity operator. My goal is to find the time-dependent state function of a SQUID prepared in state $|+\rangle$ or $|-\rangle$ at time $t=0$. Given that state function, I can grind out the conditional probabilities (and correlation coefficients) invoked below. To find that state function, I'll first derive the relevant energy eigenstates.

No other current eigenstates besides $|+\rangle$ and $|-\rangle$ contribute appreciably to the two lowest energy eigenstates. The "mixing" with higher current states is negligible, because the energies associated with $|I=\pm 2I_0\rangle$, $|I=\pm 3I_0\rangle$, etc., are very high compared to the energies associated with $|+\rangle$ and $|-\rangle$. So, to excellent approximation, the state space from which I'm going to construct the lowest-energy eigenstates ($|E_0\rangle$ and $|E_1\rangle$) is spanned by two states, $|+\rangle$ and $|-\rangle$.

Given this approximation, the usual way to derive the energy eigenstates is to invoke the symmetry of the potential to argue that the Hamiltonian takes the form $\mathbf{H}=\begin{bmatrix} E_i & b \\ b & E_i \end{bmatrix}$ in the basis $|+\rangle=\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|-\rangle=\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The Hamiltonian must be symmetric under transposition, since transposing the matrix corresponds to mirror-reflecting the potential well around $I=0$ (i.e., re-labeling $|+\rangle$ as $|-\rangle$ and vice versa). And it must

contain off-diagonal terms, or else |+> and |-> would be energy eigenstates, a conclusion we know to be wrong both phenomenologically and theoretically. Theoretically speaking, the energy eigenstates in a symmetric potential must by symmetric and antisymmetric. By graphing |+> and |-> as "spikes" on the above potential diagram, you can see that neither state is symmetric or antisymmetric. So, the "b" cross terms are needed.

It's easy to find the eigenvalues and eigenvectors of **H**. But I'm going to use a cute shortcut to find those eigenvectors (i.e., the energy eigenstates). As just mentioned, in a symmetric potential, the energy eigenstates are symmetric and antisymmetric; in other words, they're eigenstates of the parity operator **P**. Now obviously, **P**|+>=|-> and **P**|->=|+>, since one state "turns into the other" if we mirror-reflect the system. By inspection, the only way to create normalized parity eigenstates in the state space spanned by |+> and |-> is as follows:

$$|E_0\rangle = \frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)$$
$$|E_1\rangle = \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle),$$

where I've chosen arbitrary phases, and where the ground state is symmetric (**P**$|E_0\rangle=|E_0\rangle$) and the first excited state is antisymmetric (**P**$|E_1\rangle=-|E_1\rangle$). So, the current-eigenstates are superpositions of the first two energy eigenstates:

$$|+\rangle = \frac{1}{\sqrt{2}}(|E_0\rangle + |E_1\rangle)$$
$$|-\rangle = \frac{1}{\sqrt{2}}(|E_0\rangle - |E_1\rangle)$$

Therefore, if we prepare a SQUID in one of these current eigenstates, it won't stay in that eigenstate. Rather, it will oscillate back and forth between |+> and |->, just as the ammonium molecule oscillates back and forth between inversion states. See Bransden and Joachain (1989, pp. 649-653), or Cohen-Tannoudji *et. al.* (1977, pp. 464-466).

For instance, suppose the quantum state at time t=0 is $|\Psi(t=0)\rangle = |+\rangle = \frac{1}{\sqrt{2}}(|E_0\rangle + |E_1\rangle)$. Define $\omega_0=E_0/\hbar$ and $\omega_1=E_1/\hbar$. Then the SQUID's state at arbitrary later time t, assuming negligible dissipation, is

$$\begin{aligned}|\Psi(t)\rangle &= \frac{1}{\sqrt{2}}[|E_0\rangle e^{-i\omega_0 t} + |E_1\rangle e^{-i\omega_1 t}] \\ &= \frac{1}{\sqrt{2}}[\frac{1}{\sqrt{2}}(|+\rangle + |-\rangle)e^{-i\omega_0 t} + \frac{1}{\sqrt{2}}(|+\rangle - |-\rangle)e^{-i\omega_1 t}] \\ &= \frac{1}{2}[|+\rangle(e^{-i\omega_0 t} + e^{-i\omega_1 t}) + |-\rangle(e^{-i\omega_0 t} - e^{-i\omega_1 t})].\end{aligned}$$

To express this in prettier form, multiply all the exponential terms by $e^{i(\omega_0 + \omega_1)t/2}$, and pull a factor of $e^{-i(\omega_0 + \omega_1)t/2}$ out front. (So, on net, I'm multiplying the right-hand side by 1.) This "trick" yields

$$\begin{aligned}|\Psi(t)\rangle &= \frac{1}{2}e^{-i(\omega_0+\omega_1)t/2}[|+\rangle(e^{i(\omega_1-\omega_0)t/2} + e^{-i(\omega_1-\omega_0)t/2}) \\ &\qquad + |-\rangle(e^{i(\omega_1-\omega_0)t/2} - e^{-i(\omega_1-\omega_0)t/2})] \\ &= e^{-i(\omega_0+\omega_1)t/2}[|+\rangle\cos\frac{\omega}{2}t + |-\rangle i\sin\frac{\omega}{2}t],\end{aligned}$$

where I've defined $\omega=\frac{E_1 - E_0}{\hbar}$. The SQUID, just like the ammonia molecule, oscillates back and forth between its two "classical" states.

For my purposes, the relevant information to "extract" from $|\Psi(t)\rangle$ is the conditional probability of finding the SQUID in state |+> or |-> at arbitrary time t, given that the SQUID was prepared in state |+> at t=0. The calculation is trivial. In my

notation, "Q(t)=+" is shorthand for "measurement of the SQUID's current at time t yields a clockwise current."

(1) $$P_{QM}[Q(t)=+ \mid Q(t=0)=+] = |\langle +|\Psi(t)\rangle|^2 = \cos^2 \frac{\omega}{2}t$$
$$P_{QM}[Q(t)=- \mid Q(t=0)=+] = |\langle -|\Psi(t)\rangle|^2 = \sin^2 \frac{\omega}{2}t.$$

Readers familiar with Bell-type experiments may recognize these cosine-squared and sine-squared conditional probabilities. Consider the canonical Bell thought experiment, in which two spin-1/2 particles in their singlet state rush in opposite directions and undergo measurement of (perhaps different) components of spin. Suppose particle 1 is measured to have S_z=+, and suppose particle 2 undergoes measurement of S_n, where the angle between **z** and **n** is 180°-θ. Then, the quantum probability that particle 2 will yield spin up vs. spin down is

$$P_{QM}(S_n=+ \mid S_z=+) = \cos^2 \frac{\theta}{2}$$
$$P_{QM}(S_n=- \mid S_z=+) = \sin^2 \frac{\theta}{2}.$$

The spin correlations in standard Bell-type experiments are formally equivalent to the temporal correlations predicted to occur between successive current measurements on a SQUID. Therefore, quantum mechanics predicts a violation of a temporal Bell inequality, provided that the inequality is formally equivalent to a standard Bell inequality violated by spin-1/2 systems. Leggett in section 3.3, and I in section 3.4, will take advantage of this fact.

Section 3.3: Leggett's inequality: Derivation & interpretation

In this section, I'll present my version of Leggett and Garg's derivation of a temporal Bell inequality. First, I'll present their assumptions (Macrorealism and Non-invasive measurability). Then I'll show how those assumptions lead to an inequality violated by any theory consistent with QM's statistical predictions. I'll also briefly discuss how these inequalities could be tested in the lab. Finally, I'll begin my critique of Leggett's interpretation of these results, along the following lines: Leggett argues that Non-invasive measurability is a "natural corollary" of Macrorealism, and hence, any theory that violates Non-invasive measurability isn't really macrorealistic in some sense. Therefore, violation of the temporal Bell inequality implies that Macrorealism fails. In response, I'll argue that in certain kinds of theories, Macrorealism could hold even though Non-invasive Measurability fails. Therefore, even if experiments violate the temporal Bell inequalities, we can't jump to conclusions about the failure of Macrorealism. Further argument is needed to pin down the philosophical implications of such a violation. (In section 3.4, I'll pursue that project.)

§3.3.1. Leggett's conditions

Leggett and Garg's (1985) first assumption is Macrorealism:

Macrorealism: A macroscopic system with two or more macroscopically distinct states available to it will at all times *be* in one of those states.

Quantum mechanics under a standard Copenhagen interpretation violates this condition: A SQUID described by $|\Psi(t)\rangle = e^{-i(\omega_0+\omega_1)t/2}[|+\rangle \cos \frac{\omega}{2}t + |-\rangle i \sin \frac{\omega}{2}t]$

occupies a macroscopic superposition in which it doesn't actually possess either a clockwise or a counterclockwise current, just as an electron in a double-slit experiment cannot be said to *actually* traverse the left slit or the right slit. But remember, Leggett is considering whether an alternative ("hidden-variable") theory could satisfy certain natural "classical" conditions, one of which is Macrorealism. Keep in mind that Macrorealism allows superposed properties in the microscopic realm. It demands only that *macroscopic* physical quantities belonging to macroscopic objects always possess definite values.

Leggett's second assumption is

Non-invasive measurability: It is possible, in principle, to determine the (macroscopic) state of a system with arbitrarily small perturbation of its subsequent dynamics.

When we measure the SQUID's current, we can't help exerting some "back-action." This is true in both quantum and classical mechanics. But in "classical" theories, that back action can (in principle) be made arbitrarily small. In that case, the SQUID's post-measurement state evolution will (to good approximation) proceed *as if* the measurement hadn't occurred. I'll formalize and discuss this condition more fully in section 3.4 below. For now, let me show how these conditions lead to a temporal Bell inequality.

§3.3.2. Temporal Bell inequality

In a macrorealistic framework, we can let Q(t) denote the SQUID's current direction at time t. Let Q(t)=+1 and Q(t)=-1 denote clockwise and counterclockwise current,

respectively. According to Non-invasive measurability, the value of $Q(t_3)$ does not depend on whether the SQUID underwent a (sufficiently careful) earlier measurement at time t_1 or t_2, because the SQUID's state evolution proceeds as if the earlier measurement hadn't occurred. So, $Q(t_3)$ is uniquely defined. Therefore, if we measure the SQUID's current at t_1 and t_3, and obtain $Q(t_3)=+1$, then we *would* have obtained $Q(t_3)=+1$ even if the earlier measurement had occurred at t_2 instead of t_1.

Now actually, I've just assumed counterfactual definiteness, which is valid only if the SQUID's state evolves deterministically. Here's why: If the SQUID evolves stochastically, then we can't say what would have happened had the earlier measurement occurred at t_2 instead of t_1. That's not because measuring the SQUID at t_2 instead of t_1 "disturbs" the state evolution. It's simply because, if we "rerun" the SQUID's state evolution in a stochastic universe, we might get a different result, even if all initial and intervening conditions are the same. That's just what it means to be stochastic! So, my derivation of Leggett and Garg's temporal Bell inequality implicitly assumes determinism. But in section 3.4 below, I'll derive an equivalent Bell inequality in a stochastic framework. For now, I'll stick to a deterministic framework in order to keep the exposition simple.

Consider the following expression:

$$(*) \qquad Q(t_1)Q(t_3) + Q(t_1)Q(t_4) + Q(t_2)Q(t_3) - Q(t_2)Q(t_4).$$

As just noted, by Non-invasive measurability and counterfactual definiteness, the "$Q(t_3)$" paired with $Q(t_1)$ equals the $Q(t_3)$ paired with $Q(t_2)$. Each of the four Q's can equal ± 1.

This expression has only two possible values, ±2. To see this, rewrite the expression as

$$Q(t_1)[Q(t_3) + Q(t_4)] + Q(t_2)[Q(t_3) - Q(t_4)].$$

Clearly, if $[Q(t_3) + Q(t_4)]=\pm 2$, then $[Q(t_3) - Q(t_4)]=0$; and vice versa. So, the overall expression can only equal ±2.

Now suppose we consider N SQUIDs (or if you prefer, N different experimental runs on the same SQUID). Let Q_i refer to the i-th SQUID. As just shown, for each of those N SQUIDs,

$$|Q_i(t_1)Q_i(t_3) + Q_i(t_1)Q_i(t_4) + Q_i(t_2)Q_i(t_3) - Q_i(t_2)Q_i(t_4)| \leq 2. \quad (2)$$

Therefore,

$$|Q_i(t_1)Q_i(t_3) + Q_i(t_1)Q_i(t_4) + Q_i(t_2)Q_i(t_3) - Q_i(t_2)Q_i(t_4)| \leq 2.$$

From the triangle (Schwartz) inequality, a sum of absolute values is less than or equal to the absolute value of the sum: $|\sum_{i=1}^{N} b_i| \leq \sum_{i=1}^{N} |b_i|$. So, we have

$$|\frac{1}{N}\sum_{i=1}^{N} Q_i(t_1)Q_i(t_3) + \frac{1}{N}\sum_{i=1}^{N} Q_i(t_1)Q_i(t_4) + \frac{1}{N}\sum_{i=1}^{N} Q_i(t_2)Q_i(t_3) - \frac{1}{N}\sum_{i=1}^{N} Q_i(t_2)Q_i(t_4)| \leq 2.$$

In the limit as $N\rightarrow\infty$, $\frac{1}{N}\sum_{i=1}^{N} Q_i(t_1)Q_i(t_3)$ is the correlation coefficient between $Q_i(t_1)$ and $Q_i(t_3)$, by which I mean the expectation value of that joint measurement result. In the

usual notation, in the $N\rightarrow\infty$ limit, $\frac{1}{N}\sum_{i=1}^{N} Q_i(t_1)Q_i(t_3) = \langle Q_i(t_1)Q_i(t_3)\rangle$. So, the above messy inequality can be abbreviated as

$$(3) \qquad |\langle Q(t_1)Q(t_3)\rangle + \langle Q(t_1)Q(t_4)\rangle + \langle Q(t_2)Q(t_3)\rangle - \langle Q(t_2)Q(t_4)\rangle| \leq 2.$$

This inequality involving correlation coefficients is precisely what we can test via experiment. Notice that eq. (3) is formally equivalent to the Stapp-Eberhard-Redhead form of a regular Bell inequality. And recall from eq. (1) in section 3.2 that the relevant SQUID conditional probabilities (and hence the correlation coefficients) are formally equivalent to the spin-singlet state correlations. So, since spin systems violate the Stapp-Eberhard-Redhead inequality, SQUIDs violate Leggett's inequality (3), according to any theory that reproduces the statistical predictions of QM.

Let me prove this explicitly. I'll start by deriving the general expression for $\langle Q(t_a)Q(t_b)\rangle$ for a SQUID prepared in quantum state $|+\rangle$ at time t=0. It's clearly[15]

$$\begin{aligned}\langle Q(t_a)Q(t_b)\rangle \; = {} & (+1)(+1)(P_{QM}[Q(t_a)=+ \mid Q(t=0)=+])(P_{QM}[Q(t_b)=+ \mid Q(t_a)=+]) \\ & + (+1)(-1)(P_{QM}[Q(t_a)=+ \mid Q(t=0)=+])(P_{QM}[Q(t_b)=- \mid Q(t_a)=+]) \\ & + (-1)(+1)(P_{QM}[Q(t_a)=- \mid Q(t=0)=+])(P_{QM}[Q(t_b)=+ \mid Q(t_a)=-]) \\ & + (-1)(-1)(P_{QM}[Q(t_a)=- \mid Q(t=0)=+])(P_{QM}[Q(t_b)=- \mid Q(t_a)=-])\end{aligned}$$

---

[15]In the calculation, it appears that I have assumed wavefunction collapse. For the purposes of calculation, I have. But keep in mind that no-collapse QM yields the exact same conditional probabilities, as you can confirm by writing out the overall SQUID/measuring-device entangled wavefunction. Indeed, it's well known that, only by measuring certain weird holistic observables can you reveal a difference between the statistical predictions of collapse QM vs. no-collapse QM. When performing repeated measurements of the same observable, collapse and no-collapse QM always agree about conditional probabilities, correlation coefficients, and all other statistical predictions.

$$= (P_{QM}[Q(t_a)=+ \mid Q(t=0)=+])(P_{QM}[Q(t_b)=+ \mid Q(t_a)=+])$$
$$- (P_{QM}[Q(t_a)=+ \mid Q(t=0)=+])(P_{QM}[Q(t_b)=- \mid Q(t_a)=+])$$
$$- (P_{QM}[Q(t_a)=- \mid Q(t=0)=+])(P_{QM}[Q(t_b)=+ \mid Q(t_a)=-])$$
$$+ (P_{QM}[Q(t_a)=- \mid Q(t=0)=+])(P_{QM}[Q(t_b)=- \mid Q(t_a)=-])$$

$$= (\cos^2 \frac{\omega}{2}t_a)(\cos^2 \frac{\omega}{2}(t_b-t_a))$$
$$- (\cos^2 \frac{\omega}{2}t_a)(\sin^2 \frac{\omega}{2}(t_b-t_a))$$
$$- (\sin^2 \frac{\omega}{2}t_a)(\sin^2 \frac{\omega}{2}(t_b-t_a))$$
$$+ (\sin^2 \frac{\omega}{2}t_a)(\cos^2 \frac{\omega}{2}(t_b-t_a))$$

$$= (\cos^2 \frac{\omega}{2}t_a)[\cos^2 \frac{\omega}{2}(t_b-t_a) - \sin^2 \frac{\omega}{2}(t_b-t_a)]$$
$$+ (\sin^2 \frac{\omega}{2}t_a)[\cos^2 \frac{\omega}{2}(t_b-t_a) - \sin^2 \frac{\omega}{2}(t_b-t_a)]$$

$$= \cos^2 \frac{\omega}{2}(t_b-t_a) - \sin^2 \frac{\omega}{2}(t_b-t_a)$$

$$= \cos \omega(t_b-t_a),$$

where in the last two steps I used trig identities. For a surprisingly large range of choices of t_1 through t_4, inequality (3) is violated. For instance, pick $t_1=0$, $t_2=\frac{7\pi}{4\omega}$, $t_3=\frac{2\pi}{\omega}$, and $t_4=\frac{9\pi}{4\omega}$. Then, the left-hand side of inequality (3) equals

$$|\langle Q(t_1)Q(t_3)\rangle + \langle Q(t_1)Q(t_4)\rangle + \langle Q(t_2)Q(t_3)\rangle - \langle Q(t_2)Q(t_4)\rangle|$$
$$= |\cos 2\pi + \cos \frac{\pi}{4} + \cos \frac{\pi}{4} - 0|$$
$$= |1 + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} - 0|$$
$$\approx 2.41,$$

which violates the inequality.

In brief: Any SQUID theory consistent with QM's statistical predictions violates an inequality derived from Macrorealism and Non-invasive measurability.

§3.3.3. Philosophical implications: Is Macrorealism the culprit?

According to Leggett, Non-invasive measurability is a "natural corollary" to Macrorealism, by the following argument: If the SQUID actually *has* a discrete macroscopic current, then we expect the current not to get "knocked around" by a very careful measurement, since only a large disturbance could "knock" a clockwise current so hard that it becomes counterclockwise. Leggett doesn't claim that Macrorealism logically implies Non-invasive measurability. He claims only that we physically expect Macrorealistic theories "automatically" to obey Non-invasive measurability. If this argument is true, then experimental violation of Leggett's inequality would strongly suggest that nature in fact violates Macrorealism.

I'll dispute this conclusion by outlining a few classes of Macrorealistic theories that could plausibly violate Non-invasive measurability. For the remainder of this subsection, let "z" denote the parameter or parameters that control when the SQUID's current "flips" from clockwise to counterclockwise (and vice versa).

Case 1: Chaos. Suppose that the SQUID's definite current isn't always exactly $I=\pm I_0$. Instead, when the SQUID's current is clockwise, it's actually in a narrow range centered on the expected value: $I_0-\Delta I < I < I_0+\Delta I$, where the "current spread" ΔI is tiny. (Note that ΔI in this case represents an epistemic, not an ontological uncertainty.) Similarly for counterclockwise current. Suppose z is a function of I and/or microscopic degrees of freedom of the SQUID. If z follows a chaotic equation of motion, then the

back action produced by even the most careful measurement will affect the later state-evolution of the SQUID.

I should note that even in quantum theory, there might be some spread in I. Remember, I is quantized because the magnetic flux through the SQUID ring is quantized. If (for instance) the radius of the ring gets a little bigger, the value of I corresponding to one Planck-unit of magnetic flux changes slightly. Consequently, if the ring radius fluctuates even a tiny bit during the measurement interaction--due to a back-action magnetic field, for instance--ΔI will be nonzero. And if z(I) is truly chaotic, it simply doesn't matter how small ΔI is; Non-invasive measurability will be violated, if we allow the SQUID to evolve for enough time.

So, SQUID experiments cannot rule out "case 1" theories, which violate Non-invasive measurability despite the fact that they obey Macrorealism.

Case 2: Delicate flip parameter. Suppose that, even though the SQUID's current is macroscopic and definite, the "flip parameter" z depends sensitively on microscopic degrees of freedom of the SQUID. Then, even the most careful measurement could "disturb" z, and hence, the SQUID's later state evolution.

In response, Leggett could argue that a truly "classical" theory would not incorporate a "delicate z." To formalize this claim, he could argue that in a classical theory, the state evolution of macroscopic quantities depends only on the values of macroscopic quantities. For instance, in Newtonian physics, the time evolution of the center-of-mass position of a baseball depends only on the initial position, initial velocity, and net force as a function of time. Microscopic degrees of freedom simply "don't matter." Similarly, Leggett could argue, in a "classical" SQUID theory, we expect z to depend only on I (and perhaps on other macro-parameters such as the SQUID's diameter).

I think this discussion serves to show that not all Macrorealistic theories obey all of our classical intuitions. But the point still remains that an experimental violation of Leggett's inequality would not force us to renounce Macrorealism *per se*. It would force us only to renounce Macrorealistic theories that obey further classical conditions, conditions that imply (or at least strongly motivate) Non-invasive measurability.

Case 3: Bohm-type theories. In Bohm's "pilot wave" theory, particles at all times possess definite positions. A particle's trajectory is determined, in part, by a "quantum potential" corresponding to the particle's wavefunction. (Roughly put, the wavefunction is taken to be a physically real "pilot wave" that guides the particle through space.) The wavefunction obeys Schrödinger's equation. According to Schrödinger's equation, if two objects interact, their wavefunctions become entangled. In particular, a "measurement" interaction, no matter carefully performed, necessarily involves an entanglement between the system's wavefunction and the measuring device's wavefunction. (This is true even for "null-result" measurements, as I'll discuss below.) Therefore, by measuring a system, you automatically entangle its wavefunction, thereby altering the quantum potential that guides the particle's motion. So, Bohm's theory, like QM itself, violates Non-invasive measurability. (In section 3.4, I'll further discuss why QM violates Non-invasive measurability.)

Now imagine a Bohm-type theory about SQUIDs, in which the SQUID's current is always definite, and the value of the current is guided in part by the wavefunction. Such theories are Macrorealistic. But, as just shown, they violate Non-invasive measurability, due to wavefunction entanglement between the SQUID and its measuring apparatus.

These examples, and others like them, establish that physically sensible Macrorealistic theories could violate Non-invasive measurability. Therefore, assuming Leggett's inequality is violated by experiment, we should not jump to the conclusion that

Macrorealism fails. Perhaps Non-invasive measurability--and Non-invasive measurability alone--in the "culprit." I'm not claiming here that Macrorealism holds. I'm claiming only that Leggett's considerations do not *force* us to abandon Macrorealism, even if experiments violate his inequality. Indeed, in section 3.4, I'll present evidence that violation of Leggett's inequality forces us to renounce Non-invasive measurability, not Macrorealism.

Before launching into that argument, however, I'll briefly discuss the best candidate to date for a "non-invasive" measurement scheme.

§3.3.4. Testing Leggett's inequality: Null-result measurements

Tesche (1990) proposes an experiment that uses "null-result" measurements to test Leggett's inequalities against QM's statistical predictions. Her measuring device is designed to register a response if the SQUID occupies one eigenstate of current, say, clockwise. If the SQUID occupies the other current eigenstate (counterclockwise), then her device registers *no* response and exerts negligible back action on the SQUID's quantum state evolution. If we assume Macrorealism, then a null measurement result (i.e. no response) indicates, in this example, that the SQUID *has* counterclockwise current. And if that null-result measurement causes negligible back action, as we intuitively expect, then the time evolution of the SQUID's macroscopically distinct state is undisturbed. Consequently, null-result measurements are a good candidate for "revealing" the non-invasive measurability of SQUIDs, if Non-invasive measurability in fact holds.

To experimentally determine the correlation coefficients $\langle Q(t_a)Q(t_b)\rangle$ using Tesche's "trick," prepare the SQUID in the relevant initial state, $|+\rangle$. Let it evolve until time t_a, and then perform a null-result measurement. If you get the null result, say

counterclockwise, then take another measurement at time t_b. (That second measurement can be invasive, because no further measurements will occur.) If you don't get the null result at time t_a, then simply record that result as "clockwise," and don't bother taking another measurement at time t_b.

After getting good statistics, do the same experiment, but "reset" the null-result measurement device so that "clockwise" now corresponds to the null result.

At the end of all this, you'll have good statistics on what fraction of the SQUIDs were "clockwise" vs. "counterclockwise" at time t_a. So, you know $p[Q(t_a)=+ \mid Q(t=0)=+]$ and $p[Q(t_a)=- \mid Q(t=0)=+]$. You've also measured the conditional probability that a $Q(t_a)=+$ SQUID later yields $Q(t_b)=+$ vs. $Q(t_b)=-$, and the conditional probability that a $Q(t_a)=-$ SQUID later yields $Q(t_b)=+$ vs. $Q(t_b)=-$. That is, you've measured all the conditional probabilities of the form $p[Q(t_b)=\pm \mid Q(t_a)=\pm]$. So, you can immediately calculate the experimental value of $\langle Q(t_a)Q(t_b)\rangle$ using

$$\begin{aligned}\langle Q(t_a)Q(t_b)\rangle_{exp.} = {} & (+1)(+1)(p[Q(t_a)=+ \mid Q(t=0)=+])(p[Q(t_b)=+ \mid Q(t_a)=+]) \\ & + (+1)(-1)(p[Q(t_a)=+ \mid Q(t=0)=+])(p[Q(t_b)=- \mid Q(t_a)=+]) \\ & + (-1)(+1)(p[Q(t_a)=- \mid Q(t=0)=+])(p[Q(t_b)=+ \mid Q(t_a)=-]) \\ & + (-1)(-1)(p[Q(t_a)=- \mid Q(t=0)=+])(p[Q(t_b)=- \mid Q(t_a)=-]).\end{aligned}$$

Of course, this procedure assumes that nature doesn't throw us a "biased" sample of SQUIDs when the null-result measuring device is set to "clockwise" vs. "counterclockwise." For a detailed discussion of this kind of "randomness" assumption, see chapter 4 of Redhead (1987).

Section 3.4: Derivation of Leggett's inequality from weaker assumptions

In this section, we derive Leggett's inequality from assumptions weaker than Leggett's. (By "we," I mean Sara Foster and myself.) The first condition is a stochastic version of non-invasive measurability. The second condition is a realism assumption much weaker than Macrorealism, a condition obeyed by any Markovian theory, including QM. Consequently, if QM's predictions hold and therefore Leggett's inequalities fail, no Markovian theory underlying or replacing QM can allow the possibility of non-invasive measurements, even for null-result measurements on macroscopic systems.

This result clarifies the philosophical meaning of Leggett's inequality. As discussed above, Leggett's derivation shows that we must renounce Non-invasive measurability *or* Macrorealism (or both). Our new proof singles out Non-invasive measurability as the condition we must renounce. And if you buy Leggett's argument that Non-invasive measurability is a "natural corollary" to Macrorealism (despite my critique in section 3.3.3), then our new proof shows that you must renounce both Macrorealism *and* Non-invasive measurability.

§3.4.1. Notation and preliminaries

Like Leggett, we consider possible theories in which presently-unknown parameters supplement or replace the quantum state description of the SQUID. Let $\lambda(t)$ denote the SQUID's fully specified macrostate at time t. The macrostate is the aspect of the

SQUID's state causally relevant to electric current measurement results. So, $\lambda(t)$ contains all relevant information about the SQUID's macroscopic current characteristics and about the SQUID's probability of evolving into a different macrostate at a later time. According to QM, $\lambda(t)$ is just $\Psi(t)$, the quantum state vector. But in a general theory, λ may encode information not contained in Ψ. We do *not* assume that λ specifies a definite current for the SQUID; our λ states do not necessarily correspond to Leggett's "macroscopically distinct states." For instance, as in QM, λ may encode measurement-result *probabilities* instead of *certainties*.

The macrostate λ, as defined by us, does not contain information about microscopic degrees of freedom irrelevant to macroscopic current measurement results, if such degrees of freedom exist.

Let $\mu(t)$ denote the state of the device used to measure the SQUID's current, according to the general theory. In some theories, this "apparatus microstate" plays an important role.

A SQUID in quantum state Ψ could occupy one of many underlying states λ. Let $\rho[\lambda(t_1)]$ denote the probability density that a SQUID prepared to be in quantum state Ψ at time t_1 occupies state λ at t_1. Similarly, $\rho[\mu(t)]$ is the probability density that an apparatus set to measure the SQUID's current at time t occupies state μ at t. Assume

ρ-independence:

(i) $\rho[\lambda(t_1)]$ does not depend on whether the SQUID undergoes a measurement at time $t \geq t_1$.

(ii) $\rho[\mu(t)]$ does not depend on whether the SQUID underwent another measurement before t.

The first part of this condition is trivial. It requires that the SQUID not "know in advance" whether it is going to be measured at a later time. If this condition fails, then either some kind of backwards-in-time causation influences the SQUID's state preparation, or else nature has "conspired" to bias our sample of SQUIDs. This requirement is weaker than Particle Locality, the ρ-independence condition used in Bell derivations. Condition (i), unlike Particle Locality, allows spacelike influences. Condition (i) only rules out *backwards* timelike influences.

Condition (ii) demands that, if the same device measures the SQUID at t_1 and t_2, then we can "zero" (reset) the device between measurements. An easy way to ensure that this condition holds is to use different measuring devices at t_1 and t_2, and to "zero" the t_2 device after time t_1. Given this scheme, ρ-independence condition (ii) fails only if the first measuring device somehow sends information to the second device, information that "survives" when the second device gets reset. Although considerations of local causality do not rule out this possibility, it seems mighty conspiratorial and *ad hoc,* especially if we don't even turn on the second device until the first one gets shut off.

Let $M_+(t)$ denote the performance of a null-result measurement at time t such that a null response is taken to indicate positive (clockwise) current. $M_+(t)=+1$ denotes that such a measurement indeed gave the null result, signaling positive current. Define $M_-(t)$ and $M_-(t)=-1$ analogously for negative (counterclockwise) current.

As in chapter 2, I'll continue to use standard conditional probability notation. For instance, $p[M_+(t)=+1 \mid \lambda(t), \mu(t)]$ is the probability, according to the general (hidden-variable) theory, that a SQUID in state λ at t, upon measurement at t with an apparatus in state μ, would yield the null result, indicating positive current. Similarly,

$$p[M_+(t_2)=+1 \mid M_+(t_1)=+1, \lambda(t_1), \mu(t_2)]$$

is the probability that a SQUID in state λ at t_1, upon measurement at t_2 with an apparatus in state μ, would yield positive current, *given* that an earlier null-result measurement at t_1 indicated positive current. By a "measurement at t_1," we mean a measurement that begins at t_1.

$p[M_+(t_2)=+1, M_+(t_1)=+1 \mid \lambda(t_1)]$ is the joint probability that a SQUID in state λ at t_1, upon sequential null-result measurements, would yield the null result both times, indicating positive current.

§3.4.2. The main assumptions

We now discuss the two primary assumptions we'll use to derive Leggett's inequality. (Actually, we'll be deriving the temporal equivalent of the stochastic Clauser-Horne inequality, which turns out to be statistically equivalent to inequality (3) above. More on this later.) Our conditions, like Leggett's, can be tested by Tesche's null-result measurement procedure.

Non-invasive measurability for null-result measurements. In the context of general (perhaps nonmacrorealistic, perhaps stochastic) theories, a null-result measurement is

designed to register a response for only one of the two possible measurement results on the SQUID. Such a measurement, even if carefully done, may disturb some microscopic degrees of freedom of the SQUID. For instance, the back action of the measuring device might jiggle some electrons in the SQUID ring. But suppose that the SQUID's macroscopic current characteristics over time, as encoded by the macrostate, do not depend too delicately on these microscopic degrees of freedom. Then the back action only negligibly disturbs the SQUID's macrostate and its evolution. We explore this possibility by assuming

Non-invasive measurability for null-result measurements (NIMN): *The evolution of the SQUID's macrostate is disturbed arbitrarily weakly by a sufficiently careful null-result measurement (when the null result occurs).*

This is essentially Leggett's Non-invasive measurability, rephrased so as not to presuppose Macrorealism. Whether Tesche's experiment is "sufficiently careful" is, of course, an open question. NIMN demands only that such an experiment be possible in principle, even if technology hasn't reached that level.

One could argue that non-invasive measurability is intuitively compelling only for macrorealistic theories, and therefore NIMN is physically unmotivated in the more general case. We disagree. Whether or not λ specifies a definite current, NIMN will hold provided that λ does not depend too delicately on the SQUID's microstate and provided that obtaining a measurement result doesn't "automatically" collapse or "effectively collapse" the SQUID's density operator, as happens in QM. QM-style

effective collapse is not a necessary feature of all non-macrorealistic theories. It's simply a feature of the most familiar non-macrorealistic theory.

We now express NIMN mathematically. According to a general SQUID theory, a freely evolving SQUID in state λ at time t_1 has a certain *probability* (or probability density) of occupying state λ' at later time t_2. Let $\rho[\lambda'(t_2) \mid \lambda(t_1)]$ denote this state-evolution probability density. Similarly, $\rho[\lambda'(t_2) \mid \lambda(t_1), M_+(t_1)=+1]$ is the probability that a SQUID in state λ at t_1 would occupy state λ' at t_2, given that a null-result measurement beginning at t_1 indicated positive current (by null result). According to NIMN,

NIMN

$$\rho[\lambda'(t_2) \mid \lambda(t_1), M_+(t_1)=+1] = \rho[\lambda'(t_2) \mid \lambda(t_1)]$$

$$\rho[\lambda'(t_2) \mid \lambda(t_1), M_-(t_1)=-1] = \rho[\lambda'(t_2) \mid \lambda(t_1)]$$

Of course, these equations apply only when $\rho[M_+(t_1)=+1 \mid \lambda(t_1)]\neq 0$ and $\rho[M_-(t_1)=-1 \mid \lambda(t_1)]\neq 0$, respectively. In deterministic theories, all the $\rho[... \mid \lambda(t_1)]d\lambda$ are equal to zero or one.

Quantum mechanics violates NIMN. Even during an ideal measurement, null-result or otherwise, a SQUID formerly in quantum state Ψ becomes "entangled" with the measuring apparatus, and therefore the density operator describing the SQUID changes. (Certain so-called "interference terms" get smaller.) Or, if we assume wavefunction collapse, measurement collapses the SQUID into an eigenstate of current. Either way, a measurement entangles or collapses the SQUID's state, thereby changing

the SQUID's density operator, and hence, its state evolution. This is true, according to QM, *even when the null result is obtained.* So, QM violates NIMN.

SQUID Completeness. Our second major assumption is

SQUID Completeness: *A SQUID measurement-result probability at time t depends only on the SQUID's state (and on the measuring device state) at time t.*

SQUID Completeness is an incredibly weak realism assumption. It demands only that the state of the SQUID (and its measuring apparatus) *completely* determine measurement-result probabilities.[16] SQUID Completeness does not presuppose a definite current for the SQUID.

QM obeys SQUID Completeness: Given the type of measurement (i.e., the Hermitian operator corresponding to the measured observable), the quantum state at time t completely specifies measurement-result probabilities at time t. Since QM disobeys many "realism" assumptions, the fact that QM obeys SQUID Completeness suggests that our condition is very weak.

SQUID Completeness does *not* prohibit an influence by earlier measurements on later measurement results. For instance, by measuring the SQUID at t_1, we may disturb its state evolution so as to change measurement-result probabilities at t_2. This happens

[16]Jarrett's Completeness condition, discussed in chapter 2, encodes similar content. But Jarrett's condition rules out a nonlocal connection between a measurement result on particle 1 and a measurement result on particle 2. SQUID Completeness, by contrast, has nothing to do with locality, because there's only one measured system (a single SQUID) involved. So, SQUID Completeness can hold in nonlocal theories.

in QM: by measuring the SQUID at t_1, we entangle or collapse its state and thereby alter its state-evolution, leading to changed measurement-result probabilities at t_2. What SQUID Completeness disallows is a direct mysterious "influence" by the earlier measurement result on the later measurement result, an influence not propagated via the SQUID's state evolution. SQUID Completeness demands nothing more than Markovian state evolution: A complete specification of the present state of the SQUID (and its measuring device) must probabilistically "screen off" the SQUID's past, rendering the past states irrelevant. All theories that can be cast in terms of "state functions" automatically obey SQUID Completeness. Newtonian mechanics, relativistic mechanics, classical electromagnetism, quantum mechanics, and quantum field theory all obey SQUID Completeness. All hidden-variable theories that I know of also obey this condition. Frankly, it's hard to imagine a non-Markovian fundamental theory.

Since by definition, all characteristics of the SQUID causally relevant to Q-measurement results are encoded by the macrostate λ, we have

SQUID Completeness

$$p[M_+(t_2)=+1 \mid \lambda(t_2), \mu(t_2), M(t_1)=+1] = p[M_+(t_2)=+1 \mid \lambda(t_2), \mu(t_2)]$$

This condition demands that a measurement-result probability depend on the SQUID's present macrostate, *not* on how the SQUID reached its present macrostate. Again, SQUID Completeness is nothing more than a Markov requirement.

All macrorealistic theories are SQUID Complete (since the SQUID's definite current at time t determines the result of measuring I at time t),[17] but not vice versa. SQUID Completeness is much weaker, as just discussed.

§3.4.3. Derivation of Leggett's inequality

In this section, we derive Leggett's inequality from NIMN, SQUID Completeness, and the ρ-independence assumptions introduced in section 3.4.1. Specifically, we'll show that our conditions imply

Factorizability:

$$p[M_+(t_2)=+1,\, M_+(t_1)=+1 \mid \lambda(t_1)] = p[M_+(t_1)=+1 \mid \lambda(t_1)] \cdot p[M_+(t_2)=+1 \mid \lambda(t_1)].$$

Factorizability implies the Clauser-Horne version of Bell's inequality, as I'll discuss below.

To derive Factorizability, we first prove a crucial lemma.

Lemma: NIMN & SQUID Completeness & ρ-independence →

$$p[M_+(t_2)=+1 \mid \lambda(t_1)] = p[M_+(t_2)=+1 \mid M_+(t_1)=+1,\, \lambda(t_1)].$$

Proof of Lemma:

From probability theory and ρ-independence of the μ states,

---

[17]Like Leggett, I'm assuming a "Faithful Measurement" principle, according to which the value of the SQUID's definite current (when it exists) is the value "revealed" by measurement.

$$p[M_+(t_2)=+1 \mid M_+(t_1)=+1, \lambda(t_1)]$$

$$= \iint \rho[\mu(t_2)]d\mu \cdot d\lambda' \cdot \rho[\lambda'(t_2) \mid \lambda(t_1), M_+(t_1)=+1] \cdot p[M_+(t_2)=+1 \mid M_+(t_1)=+1, \lambda'(t_2), \mu(t_2)]$$

$$= \iint \rho[\mu(t_2)]d\mu \cdot d\lambda' \cdot \rho[\lambda'(t_2) \mid \lambda(t_1), M_+(t_1)=+1] \cdot p[M_+(t_2)=+1 \mid \lambda'(t_2), \mu(t_2)]$$

by SQUID Completeness

$$= \iint \rho[\mu(t_2)]d\mu \cdot d\lambda' \cdot \rho[\lambda'(t_2) \mid \lambda(t_1)] \cdot p[M_+(t_2)=+1 \mid \lambda'(t_2), \mu(t_2)]$$

by NIMN

$$= p[M_+(t_2)=+1 \mid \lambda(t_1)]$$

by probability theory.

This proves the lemma. Q.E.D.

Armed with this Lemma, we now easily prove

Theorem: NIMN & SQUID Completeness & ρ-independence → Factorizability

Proof: From probability theory,

$$(*) \qquad p[M_+(t_2)=+1, M_+(t_1)=+1 \mid \lambda(t_1)] =$$

$$p[M_+(t_1)=+1 \mid \lambda(t_1)] \cdot p[M_+(t_2)=+1 \mid M_+(t_1)=+1, \lambda(t_1)].$$

By the Lemma, the second factor on the right-hand side equals $p[M_+(t_2)=+1 \mid \lambda(t_1)]$. So, eq. (*) immediately becomes

$$p[M_+(t_2)=+1,\ M_+(t_1)=+1 \mid \lambda(t_1)] = p[M_+(t_1)=+1 \mid \lambda(t_1)]\cdot p[M_+(t_2)=+1 \mid \lambda(t_1)],$$

which is Factorizability. *Q.E.D.*

Now all that remains is to show that Factorizability, along with ρ-independence of the λ states, implies the Clauser-Horne inequalities. By reasoning equivalent to the above, our conditions also imply

$$p[M_-(t_2)=-1,\ M_+(t_1)=+1 \mid \lambda(t_1)] = p[M_+(t_1)=+1 \mid \lambda(t_1)]\cdot p[M_-(t_2)=-1 \mid \lambda(t_1)],$$

$$p[M_-(t_2)=+1,\ M_+(t_1)=-1 \mid \lambda(t_1)] = p[M_+(t_1)=-1 \mid \lambda(t_1)]\cdot p[M_-(t_2)=+1 \mid \lambda(t_1)],$$

$$p[M_-(t_2)=-1,\ M_+(t_1)=-1 \mid \lambda(t_1)] = p[M_+(t_1)=-1 \mid \lambda(t_1)]\cdot p[M_-(t_2)=-1 \mid \lambda(t_1)].$$

In other words, Factorizability holds in general, not just for specific measurement results. To get from Factorizability and ρ-independence to the Clauser-Horne inequalities takes a lot of uninstructive algebra. See Redhead (1987, chapter 4) for the boring details behind this well-known result. The outcome is a Bell inequality statistically equivalent to Leggett's inequality (3) derived above (in section 3.3.2). In other words, a theory's statistical predictions violate the Clauser-Horne inequalities if and only if they violate inequality (3) above. So, if Leggett's inequality is violated, then so is the Clauser-Horne inequality, proving that no theory about SQUIDs can obey NIMN, SQUID Completeness, and ρ-independence.

§3.4.4. Philosophical implications: Comparison to previous results

If Tesche's and others' experiments violate Leggett's inequalities, as QM predicts, then Leggett's derivation suggests that we should renounce non-invasive measurability *or* macrorealism. Since both of these assumptions are "controversial," and since a theory could obey one but not the other (as argued above in section 3.3.3), a reasonable theory could disobey either (or both!).

Our contribution is to show which assumption is probably "at fault" if QM's predictions turn out to be correct. SQUID Completeness, unlike Macrorealism, is so weak that we expect any reasonable theory to obey it. And ρ-independence only rules out conspiratorial theories, or theories that allow backwards-in-time causation. NIMN, SQUID Completeness and ρ-independence lead to a Bell-type inequality violated by any theory that reproduces QM's statistical predictions. Therefore, if QM's predictions are correct we should renounce the possibility of performing non-invasive measurements even in principle, even if we use ingenious null-result measuring procedures to measure macroscopic quantities. In brief, our derivation strongly suggests that if Leggett's inequality fails, non-invasive measurability is "to blame."

This result improves upon Ballentine (1987). Ballentine argues that non-invasive measurability alone implies Leggett's inequality. According to him, NIMN entails that the correlations between sequential SQUID measurement results do not depend on whether an intervening (non-invasive) measurement occurs. But this independence follows only if we make some assumption about the relationship between the SQUID's state and the SQUID's measurement-result probabilities. Our SQUID Completeness assumption fills precisely this gap in Ballentine's reasoning. Without SQUID

Completeness (or a stronger assumption such as macrorealism), non-invasive measurability has no empirical consequences.

If you believe (following Leggett) that no reasonable Macrorealistic theory would violate NIMN, then our result establishes that you must renounce both NIMN *and* Macrorealism, instead of one or the other.

§3.4.4. Philosophical implications: Holism.

We now see that if QM's predictions hold, and if ρ-independence and the weak form of realism encoded by SQUID Completeness hold, then failure of non-invasive measurability is not simply a quirk of the quantum formalism. Instead, that failure indicates nature's unwillingness to allow non-invasive measurements even in principle. This is what SQUIDs have to tell us about metaphysics, even though they can't tell us about Macrorealism *per se*.

But what does violation of non-invasive measurability tell us about nature? In other words *why* does non-invasive measurability fail? One possibility is that the measuring device "disturbs" the measured system significantly, in the usual "causal" sense of "disturbance." But if this were the case, then we'd expect a null-result measurement to disturb the system less than a regular measurement, when the null result is obtained. In other words, we'd expect the size of the disturbance to depend in some way on the severity of the "intrusion." But according to QM, the disturbance doesn't scale down with the intrusion in this way; even a "perfect" null-result measurement effectively collapses the SQUID's density operator, leading to a violation of Leggett's inequality. So, if we want to retain a causal picture of the measurement as an intrusion leading to a

disturbance, we have to abandon our classical causal intuitions about how the size of the cause relates to the size of the effect.

The quantum formalism suggests another metaphysical interpretation of why non-invasive measurability fails. As discussed above, any measurement, simply by virtue of being an interaction between two quantum systems, inevitably leads to wavefunction entanglement between the measuring device and measured system. After this entanglement, the two systems are holistically connected, in the following senses: Neither the SQUID nor its measuring apparatus alone *has* its own state vector. And the two-part system as a whole has properties and/or propensities that don't supervene on the separate properties/propensities of the individual systems. Put roughly, the properties of the whole don't reduce to composite properties of the parts. These holistic properties include correlations between, say, the SQUID's current and the measuring device's pointer reading. So, non-invasive measurability fails not because of some "causal" disturbance, but because the SQUID becomes holistically entangled with another system, an entanglement that changes the probabilities associated with the SQUID alone.

This holistic view of violation of non-invasive measurability helps us explain why making the measurement less "disturbing" doesn't lead to a smaller violation of non-invasive measurability (assuming QM's predictions hold). In my holistic framework, the severity of interaction leading to holistic entanglement shouldn't matter. All that matters is *whether* the systems become holistically connected. Since a measurement is an interaction *designed* to bring about a correlation between the measured system and the measuring device's "pointer reading," and since (in this framework) the correlations

result from holistic entanglement, it follows that any measurement worth its name leads to holistic entanglement.

This argument alone isn't strong enough to make you abandon your "causal" world view in favor of a "holistic" one, whatever that means. But this argument *isn't* alone. In chapter 5, I'll argue in detail that the best way to explain the local causality violations discussed *ad nauseam* in chapter 2 is to renounce "causality" in favor of "holism." And in chapter 4, I'll show how an explicitly holistic interpretation of QM may be able to account for the macroscopic world as we observe it. So, every chapter of this dissertation adds to the argument that quantum reality is best interpreted within a holistic, noncausal metaphysical framework.

CHAPTER 4: DECOHERENCE AND 'MODAL' INTERPRETATIONS OF QM

In the past five years, "decoherence" has received loads of attention. Various decoherence-based interpretations of QM claim to recover a "classical" world at the macroscopic level.[18] In this chapter, I'll critically evaluate these claims. Decoherence, I'll argue, does not *in itself* define or even a suggest an interpretation, nor does it provide us with a new metaphysical framework. It turns out, however, that results from "decoherence theory" can save certain interpretations from otherwise-fatal technical problems. Specifically, the phenomenon of decoherence can help to "pick out" a special pointer-reading basis. (But a preferred basis alone does not an interpretation make.)

To focus my analysis of what decoherence can and cannot accomplish, I'll devote substantial discussion to a promising, comparatively new class of interpretations called "modal" interpretations. After briefly outlining how these interpretations work (section 4.1), I'll show why, without decoherence, these interpretations are doomed to failure. Roughly put, the modal interpretations without decoherence pick out the "wrong" pointer-reading basis after a non-ideal measurement; and *all* measurements of certain observables are in fact non-ideal. Then, in section 4.3, I'll show how decoherence can perhaps rescue the modal interpretations (and certain other interpretations) from the "imperfect measurement problem" just mentioned. Finally, in section 4.4, I'll explore whether the modal interpretations, aided by decoherence, have a fighting chance of "solving" the measurement problem, even when the observer's brain is taken into account and treated as another quantum mechanical system. We'll see that the modal interpretations fare surprisingly well.

---

[18]The SQUIDs discussed in chapter 3 escape these claims because they interact minimally with their environment, and hence take a long time to "decohere."

Section 4.1: Modal interpretations

In this section, I'll motivate and describe the modal interpretations, and I'll show how they apparently solve the "measurement problem" in an elegant, powerful way.

§4.1.1. Historical and philosophical motivation

Instead of diving right into the formal details, let me situate and motivate the modal interpretations. Classical intuitions suggest that physical observables have definite values at all times, and hence, a probabilistic descriptions of those quantities reflects our ignorance about the actual values. But we know from Bell (1966), Kochen and Specker (1967), and similar results that we can't consistently assign values to all observables in a way consistent with both the QM formalism and certain intuitive rules. Only *some* observables may possess (noncontextual) values at a given time. But which ones? Interpretations split into two broad classes based on their answer to this question. The wedge is provided by the

Eigenvector-eigenvalue link: A physical quantity Q *has* a definite value if and only if the quantum state is an eigenstate of the corresponding Hermitian operator **Q**.

Standard Copenhagen interpretations with wavefunction collapse obey this "orthodox" value-assignment rule. In a sense, so do relative-state interpretations, according to which Q has a definite value with respect to a given branch of the superposition only if that branch is an eigenstate of **Q**. But many interpretations violate the eigenvector-eigenvalue link. Such interpretations have two choices. They can either (i) *a priori* set in stone which observables have definite values, or else (ii) let the

quantum dynamics "pick out" which observables have definite values at a given time. Bohm's interpretation is type (i): It assigns definite positions to all particles at all times, whether or not the particle occupies an eigenstate of the position operator.

By contrast, other theorists who renounce the eigenvector-eigenvalue link prefer option (ii). They prefer not to put in "by hand" which observables are definite. They think the quantum dynamics itself should select the preferred observables (i.e., the preferred basis). For instance, Zurek (1993a,b) bases an interpretation around the claim that the definite-valued observables associated with a system S are those corresponding to operators that commute with $\mathbf{H}_{int}$, the interaction Hamiltonian between S and its environment. More on this later. The modal theorists, on the other hand, claim that the quantum state picks out which observables take on definite values; and a stochastic equation of motion describes the evolution of those definite values.[19]

Why might someone prefer such a theory to Bohm's? It has to do with how far you're willing to depart from the "pure" quantum formalism. In Bohm's theory, the Hilbert space formalism describes nothing more than how a "quantum potential" (pilot wave) evolves in time. The particles *themselves* are separately real entities that follow an independent equation of motion. And the "privileged" status of position is put in by hand. By contrast, in modal interpretations, the Hilbert-space state vector describes the particles themselves, although the description it provides isn't complete. But the quantum state picks out which observables receive definite values, values that "complete" the state description. For these reasons, the modal interpretations allegedly stay "closer" to the quantum formalism than Bohm-type theories do.

§4.1.2. Modal interpretations and the measurement problem

[19]To date, a completely successful equation of motion for these "hidden variables" has not been formulated. Dieks and his group at Utrecht are working on this.

But of course, staying close to the quantum formalism is no virtue if these interpretations can't solve the measurement problem. In this section, I'll present the rule by which the modal interpretation picks out definite-valued observables in violation of the eigenvector-eigenvalue link. Throughout this chapter, my description of "the modal interpretation" will refer to the common elements shared by Dieks (1989, 1994) and Healey (1989). Those two interpretations disagree about certain subtleties that aren't relevant here. (Most of the following discussion applies also to the original modal interpretation of van Fraassen (1979), as well as the later interpretations of Kochen (1985) and Clifton (1994)). Then, I'll show how this interpretation apparently solves the measurement problem.

Modal interpretation. For simplicity, consider an isolated quantum system composed of two entangled subsystems, 1 and 2. If we let $|Q_i\rangle$ and $|R_i\rangle$ denote a complete basis for subsystems 1 and 2, respectively, then the quantum state takes the following form:

$$|\phi\rangle = \sum_i \sum_j c_{ij} |Q_i\rangle \otimes |R_j\rangle$$

The system as a whole possesses no properties other than the ones corresponding to the quantum state vector. But the individual subsystems *do* possess definite values for certain observables. To specify *which* observables, the modal interpretation takes advantage of the

Biorthogonal decomposition theorem: For any quantum state $|\phi\rangle$ describing two subsystems, there exists locally maximal Hermitian operators **A** (describing subsystem 1) and **B** (describing subsystem 2) such that $|\phi\rangle$ can be "biorthogonally decomposed" as

follows: $|\phi\rangle = \sum_i c_i|A_i\rangle\otimes|B_i\rangle$, where $\{|A_i\rangle\}$ and $\{|B_i\rangle\}$ are eigenstates of **A** and **B**. Furthermore, if all the nonzero $|c_i|$'s are distinct (i.e., if the "contributing" $|A_i\rangle$'s and $|B_i\rangle$'s are nondegenerate), then the biorthogonal decomposition is unique.

This theorem provides the modal interpretation with the "preferred basis" it needs. According to the interpretation, if $\sum_i c_i|A_i\rangle\otimes|B_i\rangle$ is the unique biorthogonal decomposition of $|\phi\rangle$ with respect to subsystems 1 and 2, then observables A and B both *have* definite (but in general unknown) values. If the biorthogonal decomposition of $|\phi\rangle$ isn't unique, then certain degenerate observables take on definite values, but no locally maximal observables do. Of course, as the quantum state evolves in time, the observables picked out by the biorthogonal decomposition keep changing. So, unlike Bohm's theory, in which position is always definite, the modal interpretation allows different observables to be "preferred" at different times.

The modal theory, we see, gives us a prescription to figure out the definite-valued observable associated with any object. Call that object subsystem 1, and the rest of the universe subsystem 2. Biorthogonally decompose the quantum state of the universe with respect to subsystems 1 and 2, and read off the observable picked out for subsystem 1. It's easy to show that the "basis" selected in this way is the basis that diagonalizes the reduced density operator describing subsystem 1. Specifically, when $|\phi\rangle = \sum_i c_i|A_i\rangle\otimes|B_i\rangle$, the reduced density operator describing subsystem 1 is $\rho=\sum_i|c_i|^2|A_i\rangle\langle A_i|$, a "mixture" of A-eigenstates. So, as van Fraassen (1991) points out, the modal interpretation assigns definite values *as if* the ignorance interpretation of mixtures were correct. According to the ignorance interpretation of mixtures, a system described by a mixture *really does* occupy one of the eigenstates in that mixture. But there's a big difference between a "true" ignorance interpretation and the modal

interpretation. As I'll discuss later, a true ignorance interpretation applies consistently only to collapse theories, in which the quantum state of subsystem 1 "collapses onto" an A-eigenstate (in this example). In the modal interpretation, by contrast, *no collapse of the wavefunction happens*. The quantum state of the universe continues evolving according to Schrödinger's equation. When the density operator of subsystem 1 is $\rho=\sum_i|c_i|^2|A_i><A_i|$, then subsystem 1 *has* a definite value for A, even though the quantum state of the universe ($|\phi> = \sum_i c_i|A_i>\otimes|B_i>$) is *not* an eigenstate of A. This reminds us that the modal interpretation violates the eigenvector-eigenvalue link. In a sense, the definite values picked out by the biorthogonal decomposition are "hidden variables," though modal interpreters resist this term. But the interpretation assigns definite values to the observables you'd "expect" by looking at density operators. In this way, modal interpreters stay "close" to the quantum formalism.

The measurement problem. Let's see how this interpretation addresses the measurement problem, which I'll now briefly review.

Consider a spin-1/2 particle initially described by a superposition of eigenstates of S_z, the z-component of spin:

$$|\varphi> = c_1|S_z=+> + c_2|S_z=->.$$

Let $|R=+>$ and $|R=->$ denote the "up" and "down" pointer-reading eigenstates of an apparatus that measures S_z. According to pure QM (with no collapse), if the apparatus ideally measures the particle, the combined system evolves into

(1) **Ideal measurement** $|\phi> = c_1|S_z=+>\otimes|R=+> + c_2|S_z=->\otimes|R=->.$

Common sense based on everyday experience insists that, after the measurement, the pointer reading is definite. But according to the eigenvector-eigenvalue link, the pointer reading is definite only if the quantum state is an eigenstate of **R**, the pointer-reading operator. Since $|\phi>$ is not an **R**-eigenstate, the pointer reading is indefinite, according to "orthodox" interpretations with no collapse. But notice that state (1) is a biorthogonal decomposition! Therefore, according to the modal interpretation, the particle *has* a definite z-component of spin, and the pointer *has* a definite reading, assuming $|c_1|\neq|c_2|$. So, the modal interpretation neatly solves the measurement problem, at least for ideal measurements. (In section 4.2 below, I'll discuss what difficulties arise for imperfect measurements.) And it does so without proposing a modification to Schrödinger's equation, such as wavefunction collapse.

A critic could object that assigning definite values based on biorthogonal decompositions (or equivalently, diagonal density operators) is an arbitrary, physically unmotivated "trick." In response, Clifton (1995) shows that, if we want the quantum state to "choose" which observables take on definite values, then the biorthogonal decomposition is the only "basis selection rule" that obeys certain natural classical conditions. But even if the modal basis selection rule weren't *a priori* unique in some sense, we'd still have to take it seriously if it solved the measurement problem.

Section 4.2: Imperfect measurements in the modal interpretation

In this section, I'll show that modal interpretations do not in fact solve the measurement problem (without the "help" of decoherence). The argument runs as follows: When the measurement interaction isn't ideal, the biorthogonal decomposition picks out a basis that might not even be "close" to the pointer-reading basis. This is relevant, because real-life measurements of some observables are *necessarily* non-ideal, according to the QM formalism.

§4.2.1. The problem with non-ideal measurements

I'll now expand upon an argument first presented by Albert and Loewer (1990) about why the modal interpretation fares poorly if measurements are non-ideal.

For concreteness, continue to consider a spin-1/2 particle about to be measured by an S_Z-measuring device. If the measurement interaction is non-ideal, then an initially spin-up ($|S_z=+\rangle$) particle has nonzero probability of yielding a "down" pointer reading ($|R=-\rangle$). Similarly, an initially spin-down particle has a nonzero probability of yielding an "up" pointer reading. Let's assume that when an initially spin-up particle yields a "down" measurement outcome, the particle's state is not always flipped into the $|S_z=-\rangle$ state. It follows from the linearity of Schrödinger's equation that the post-measurement state of the particle/apparatus system is

(2) **Imperfect measurement**

$$|\phi'\rangle = c_{11}|S_z=+\rangle\otimes|R=+\rangle + c_{12}|S_z=+\rangle\otimes|R=-\rangle + c_{21}|S_z=-\rangle\otimes|R=+\rangle + c_{22}|S_z=-\rangle\otimes|R=-\rangle,$$

where the "mistake-term" coefficients c_{12} and c_{21} are small but nonzero. Later, I'll argue that eq. (2) describes real-life measurement interactions. For now, let me assume this is the case. Notice that eq. (2) is *not* a biorthogonal decomposition. Therefore, according to the modal interpretation, neither the apparatus's pointer reading nor the particle's z-component of spin has a definite value. To find out which observables *do* have definite values, we must re-express state $|\phi'\rangle$ in eq. (2) as a biorthogonal decomposition. (Such a decomposition always exists, as noted above.) Doing so yields

$$(2') \qquad |\phi'\rangle = \sum_i d_i |S'=s_i\rangle \otimes |R'=r_i\rangle,$$

where $\{|S'=s_i\rangle\}$ are eigenstates of some operator **S'** that doesn't commute with S_z, and $\{|R'=r_i\rangle\}$ are eigenstates of some operator **R'** that doesn't commute with **R**. Physically, S' is a spin-component of the particle along some direction other than the z-direction; and R' is an observable whose eigenstates correspond to a macroscopic superposition of different pointer-readings. According to the modal interpretation, S' and R' *have* definite values, while the pointer reading does *not* have a definite value.

As I'll discuss later, this wouldn't necessarily be disastrous if R' were some observable very "close" to the pointer reading R, i.e., if the R' eigenstates were very nearly R eigenstates.[20] But, as Albert and Loewer point out, no matter what measurement-interaction Hamiltonian is assumed, there exist a range of coefficients c_1 and c_2 such that a particle initially in state $|\varphi\rangle = c_1|S_z=+\rangle + c_2|S_z=-\rangle$, upon interacting with the measuring device, results in a particle/apparatus state whose biorthogonal decomposition picks out an apparatus observable not even close to the pointer reading. As Dickson (1994) shows, the range over which c_1 and c_2 "misbehave" might be very

[20]Formally, **R'** is "close" to **R** if and only if, for all i, $\langle R'=r'_i \mid R=r_i\rangle \approx 1$.

small. But it's not clear that an evil scientist could not prepare a bunch of particles in a "misbehaving" state. In brief: A non-ideal measurement does not always yield a definite result. Therefore, if real-life measurements are indeed imperfect (as described by eq. (2)), the modal interpretation does not solve the measurement problem.

As mentioned above, eq. (2) fails to describe a non-ideal measurement only if a spin-up particle, when it mistakenly yields a "down" measurement outcomes, *always* gets its spin flipped into the $|S_z=->$ state; and vice versa. By playing around with measurement-interaction Hamiltonians, you can confirm that such 100%-reliable spin-flipping is extremely unlikely to occur. Which isn't surprising, because we have no physical reason to expect that it *would* occur.

§4.2.2. Why measurements are non-ideal, part 1

How important is the "imperfect measurement problem" just discussed? In the following two subsections, I'll prove that measurements are *always* non-ideal, according to the QM formalism itself. In other words, ideal measurements are physically impossible. It follows that no amount of technological prowess can produce an ideal measuring device, even in principle.

Here, I present a plausibility argument that measuring devices inevitably make mistakes, due to unavoidable "fluctuation" interactions between the particle/apparatus system and its environment.

Consider the following experiment: Spin-1/2 particles get shot between Stern-Gerlach magnets. A large distance behind the magnets, we place two "photographic" plates. Plate 1 lies in the "up" path of the particles, while plate 2 lies in the "down" path.

To be ideal, this measurement of S_z must satisfy the following condition: When the initial spin state of the particle is $|\phi\rangle = c_1|S_z=+\rangle + c_2|S_z=-\rangle$, then the final state of the system is

$$|\psi\rangle = c_1|\text{particle on plate 1}\rangle\otimes|\text{dot on plate 1}\rangle + c_2|\text{particle on plate 2}\rangle\otimes|\text{dot on plate 2}\rangle;$$

or, if the system becomes entangled with environmental degrees of freedom, the reduced density operator describing the particle/apparatus system must be a mixture of |particle on plate 1>⊗|dot on plate 1> and |particle on plate 2>⊗|dot on plate 2>.

Suppose an $|S_z=+\rangle$ particle passes through the magnets. A stray photon or other stray particle hitting plate 2 might initiate reactions that produce a dot, thereby registering an incorrect "down" reading. In addition, an environmental interaction might prevent the "up" dot from forming on plate 1. A photon, for instance, occasionally causes a bound electron on the surface of the plate to ionize, via the photoelectric effect. That ionized electron might "bump into" the incoming particle, preventing it from reaching plate 1.

I must stress the physical impossibility of completely eliminating these environmentally induced errors. We can cool down an experiment to reduce thermal fluctuations, but we can never reach absolute zero, even in principle. Although we can shield the experiment from electromagnetic radiation, some blackbody radiation invades even the coldest experiments, and blackbody radiation contains *all* frequencies. Under optimal conditions, environmentally induced "fluctuation" errors will occur rarely, perhaps only 10^{-1000} percent of the time. But if such mistakes have any nonzero chance of occurring, the pointer reading does not become perfectly correlated with the particle's

z-component of spin, and hence, the modal interpretation does not assign a definite value to the pointer reading.

In this section, I have not formally proven that all measurements suffer from environmentally induced errors. But I've made this assertion highly plausible.

A modal interpreter could respond as follows: For a particle described by state $|\phi\rangle = c_1|S_z=+\rangle + c_2|S_z=-\rangle$, $|c_1|^2$ specifies the probability that the particle, upon interacting with an ideal measuring device, would acquire definite value "up" for S_z. The physical impossibility of performing the relevant ideal measurement in no way threatens the coherence or beauty of this modal interpretation. To see why, consider Newton's first law. No real-life particle travels uniformly (i.e., at constant velocity in a straight line), because of the gravitational forces exerted by other particles. Nonetheless, the first law occupies a crucial place within the logical structure of Newtonian mechanics. Furthermore, in the limit as the forces acting on the particle become arbitrarily weak, the particle follows a trajectory that *approaches* a straight line. For these reasons, the physical impossibility of uniform motion in no way threatens the coherence or beauty of Newton's first law within the framework of Newtonian mechanics. Similarly, a modal interpreter could argue, the fact that measurements can only *approach* "idealness" does not threaten the coherence or beauty of the modal interpretation, even though such an interpretation rests, in part, on the notion of ideal measurement.

In reply, I would emphasize that an adequate solution to the measurement problem must explain why real-life measuring devices register (or at least seem to register) definite results. If an interpretation cannot explain why all (or at least, almost all) measuring devices appear to display definite readings, then the interpretation cannot explain our experiences. And if an interpretation can't explain our experiences, then it's inadequate, end of story. An interpretation must do more than explain the experiences of

conscious observers in a hypothetical idealized universe; it must explain *our* experiences in *our* universe. If an interpretation works only in idealized cases, then it is at best a tentative first step.

In reply, a modal interpreter could say, "Fine, my interpretation is just a first step towards a deeper understanding of quantum mechanics. But it's a *good* first step." Let me pursue this line of thought. Imagine an alternate universe, subject to Newton's laws, that contains only one particle. This particle would travel uniformly. The physical impossibility of uniform motion in a many-particle Newtonian universe follows *not* from Newtonian mechanics alone, but from the existence of many particles (along with Newtonian mechanics). Therefore, the concept of uniform motion is coherent within a Newtonian framework, even though such motion never occurs in our universe.

Similarly, a modal interpreter could argue, the faultiness of real-life measurements follows *not* from QM alone, but from the existence of certain kinds of environmental interactions (along with QM). We can imagine an alternate universe, subject to nonrelativistic quantum mechanics, that contains only two objects, a particle and a measuring apparatus. Since this fictional universe contains no stray particles, environmentally induced errors won't plague the measurement interaction between the particle and the device. My arguments so far give us no reason to deny that the measurement interaction could be ideal. Perhaps the concept of ideal measurement can coherently occupy a central place within an interpretation of QM, in which case the modal interpretation seems to be a useful "first step" toward a deeper understanding.

In the next subsection, however, I show that modal theorists cannot invoke the "good first step" argument of the previous paragraph. Specifically, I demonstrate that the quantum formalism itself rules out ideal measurements of most observables.

Therefore, an interpretation that "works" only for ideal measurements does not work at all.

§4.2.3. Why measurements are non-ideal, part 2

I now present a plausibility argument for the following claim: The faultiness of most measurements follows not just from environmental "fluctuation" interactions, but from the logical structure of QM itself.

Consider the Stern-Gerlach experiment described above. Suppose the particle initially occupies state $|S_z=+>$. This particle is described by a reasonably localized spatial wavepacket that deflects upward upon passing between the Stern-Gerlach magnets. But the spatial wavepacket has infinitely long "tails," by which I mean the wavefunction has nonzero amplitude arbitrarily far from its peak. Schrödinger's equation implies that these tails exist, *no matter what spatial wavefunction initially described the particle.* A wavefunction localized within a bounded volume at time t=0 develops infinite tails for any time t>0, except perhaps for a finite number of later times (out of the continuous infinity of times available). And a particle that begins with infinite tails keeps them forever (except perhaps at a finite number of times). Therefore, an initially $|S_z=+>$ particle has nonzero probability of "hitting" plate 2. By similar reasoning, an initially $|S_z=->$ particle has nonzero probability of producing a dot on plate 1. Consequently, this measurement is non-ideal, no matter how carefully we forge our magnets and coat our plates. The measurement error results not from environmental interactions, but from wavefunction tails. QM itself, specifically Schrödinger's equation, implies that these tails exist.[21] Therefore, QM implies that this measurement scheme

[21]Furthermore, these infinite wavefunction tails don't go away when we switch to a relativistic framework. As Fleming (1965), Ruijsenaars (1981),and Hegerfeldt (1974, 1985) show, if a Klein-Gordon or Dirac particle is localized at t=0 within a bounded volume, then the particle has nonzero probability of being found arbitrarily far away at

cannot be made ideal, even in principle, even in an alternate universe containing no stray particles.

From this example, you can see that all "indirect" measurements are intrinsically non-ideal. Measurement of A is "indirect" if, during the measurement interaction, (i) A becomes correlated with X, where X is the position of the particle (or the position of an auxiliary system), and (ii) the "pointer reading" becomes correlated with X. I conjecture that many physical quantities can only be measured indirectly. Any purported counter-example will have to withstand intense scrutiny, during which we must treat all inner workings of the measuring apparatus quantum mechanically.

Let me summarize the plausibility argument given above. For most observables, "measurement" involves the measured observable becoming correlated with the position of the particle or the position of an auxiliary system (e.g., something inside the measuring device). QM implies that the spatial wavefunction describing the particle (or the auxiliary system) has infinite tails. Consequently, when the pointer reading becomes correlated with the position of the particle (or the position of the auxiliary system), a huge "mistake" has a nonzero chance of occurring. Therefore, QM implies that indirect measurements are necessarily non-ideal. Therefore, since we have no reason to believe that all (or even most) observables can be measured directly, an interpreter of QM must not lean too heavily on the notion of ideal measurement.

In subsections 4.2.2 and 4.2.3, I've extended and clarified Albert and Loewer's arguments about imperfect measurements, by pinpointing two reasons why measuring devices make mistakes. If faulty measurements resulted solely from environmental interactions, then we could coherently ground an interpretation of QM partly on the

---

any time $t>0$; the position probability amplitude "leaks out" of the relevant light cone. But the problem remains even if the tails merely fill the forward light cone, because that's enough to ensure that an up-deflected wavepacket has nonzero probability density at the down plate.

notion of ideal measurement, invoked counterfactually. But QM itself implies that indirect measurements are non-ideal, due to wavefunction tails. In *any* universe that obeys QM, ideal indirect measurements are *physically* impossible. Therefore, if the modal interpretation works well only for ideal measurements, it is not even a "good first step" toward a deeper understanding of QM.

Given all this, things look bad for the modal interpretation. In sections 4.3 and 4.4, however, we'll see that decoherence comes to the rescue.

Section 4.3: Decoherence as savior? What decoherence can and cannot do for interpretations of QM.

In this section, I'll explore to what extent decoherence can save the modal interpretation, and also relative-state interpretations, from objections raised against them. Decoherence cannot help these interpretations address the general metaphysical challenges raised against them.[22] But decoherence can help pick out a "special" basis that determines which observables receive definite values. I'll explore to what extent decoherence rescues the modal (biorthogonal) basis-selection rule, and Zurek's (environmental interaction) basis-selection rule, from the basis degeneracy problem and the imperfect measurement problem. "Basis degeneracy" occurs when a selection rule does not pick out a unique basis. The "imperfect measurement" problem, discussed in detail in section 4.2, occurs when a selection rule, designed to choose the pointer-reading basis after an ideal measurement, chooses a basis not even close to the pointer-reading basis after a non-ideal measurement. Decoherence, we'll see, gives the modal interpretation a fighting chance of escaping these technical difficulties.

§4.3.1. The formalism of decoherence, and what it means physically

---

[22]As Arntzenius (1988) discusses, some versions of the modal interpretation--notably Kochen's and Dieks'--do not in general allow "property composition." (Healey's interpretation avoids this problem.) For instance, imagine a rock floating through space. According to the modal interpretation, it's possible that the left side of the rock has a definite position, as does the right side, even though the rock as a whole does *not* have a definite position. We'll see later the decoherence can't eliminate this metaphysical weirdness, though decoherence can assure that this weirdness almost never applies to the *position* of parts vs. wholes.

In this section, I clarify the meaning of "decoherence." This is necessary, because many claims circulating around the physics community seem to be based on a misunderstanding of what decoherence *is*.

First, let me clarify what decoherence is *not*. I'll draw an analogy with a more familiar phenomenon, friction in Newtonian mechanics. By friction, I mean regular sliding friction as well as air resistance and all "dissipative" interactions of that sort, interactions that tend to slow down and heat up macroscopic objects. Friction is not an interpretation of classical physics. Nor is it a physical phenomenon implied by the logical structure of classical physics. You can imagine a Newtonian universe consisting of just a few particles. In that universe, "friction" doesn't exist.

Instead, friction is a physical phenomenon whose existence in *our* universe is implied by classical physics. Because our universe is filled with stray particles, a macroscopic object moving through the atmosphere--or even through outer space--inevitably slows down due to Newtonian interactions with those stray particles. So, we can't "turn off" friction, at least, not completely. In some cases, it acts quickly and thoroughly, while in other cases, it can be neglected. Crucially, friction *must* be taken into account in order to explain certain phenomenon.

Decoherence plays a similar role within the realm of quantum physics. Decoherence is *not* an interpretation. Rather, it's a physical phenomenon that results from the interaction of a (usually but not necessarily) macroscopic object with many "stray particles" in its environment.[23] Decoherence is not implied by the logical structure of QM; in a quantum universe containing only one or two particles, decoherence wouldn't exist. But since our particular universe is filled with stray particles that interact with objects according to certain interaction Hamiltonians, decoherence in our universe is

[23]An object's own internal degrees of freedom interacting with each other in a "dissipative" manner can also constitute decoherence.

implied by the quantum formalism. Macroscopic objects inevitably undergo a "dissipative" interaction with their environment, i.e., an interaction that tends to wipe out the phase coherence between macroscopically distinct states. This is decoherence.

Let illustrate decoherence with an example. Consider a standard double-slit experiment, in which a bunch of "coherent" small particles pass through a double slit with a "photographic" plate behind it. After passing through the slits, a given particle is described by a macroscopic superposition, such as $|\psi\rangle = 2^{-1/2}\{|\text{passed through slit 1}\rangle + |\text{passed through slit 2}\rangle\}$. Because this "combination" of states is a superposition (instead of a "mixture"), those two states "interfere," producing the characteristic pattern on the photographic plate. But if the particle, soon after traversing the slits, interacts strongly with stray particles, it becomes entangled with those particles. Depending on the "severity" of this environmental interaction, the interference effects get more and more "washed out."

To see how this decoherence works formally, let me streamline my notation for the particle states by letting $|\phi_1\rangle$ and $|\phi_2\rangle$ denote $|\text{passed through slit 1}\rangle$ and $|\text{passed through slit 2}\rangle$. If the particle doesn't interact with its environment, then it is described by the quantum state $|\psi\rangle = 2^{-1/2}\{|\phi_1\rangle + |\phi_2\rangle\}$, corresponding to density operator

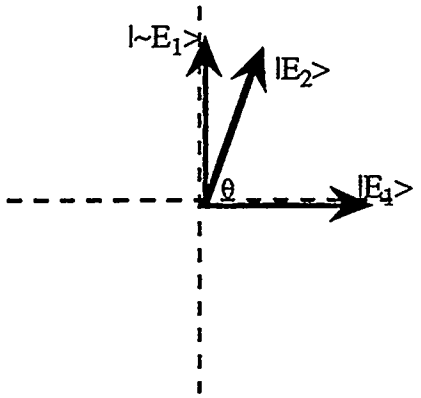
$$\rho_s = |\psi\rangle\langle\psi| = |c_{11}|^2|\phi_1\rangle\langle\phi_1| + |c_{22}|^2|\phi_2\rangle\langle\phi_2| + c_1c_2^*|\phi_1\rangle\langle\phi_2| + c_2c_1^*|\phi_2\rangle\langle\phi_1|,$$

with $c_1=c_2=2^{-1/2}$ in this case. (The subscript "s" stands for "superposition.") The "interference terms" $c_1c_2^*|\phi_1\rangle\langle\phi_2|$ and $c_2c_1^*|\phi_2\rangle\langle\phi_1|$ encode the size of the interference effects.

But if the particle interacts with its environment, it becomes entangled with environmental degrees of freedom. Let $|E_1\rangle$ and $|E_2\rangle$ denote the state of the environment after interacting with a particle in state $|\phi_1\rangle$ and $|\phi_2\rangle$, respectively. Then, the particle/environment state right before the particle reaches the photographic plate is

$$|\psi\rangle = 2^{-1/2}\{|\phi_1\rangle\otimes|E_1\rangle + |\phi_2\rangle\otimes|E_2\rangle\}.$$

We can find the new density operator describing the particle by tracing over the environmental degrees of freedom. To do so, let $|\sim E_1\rangle$ denote an environmental state that's orthogonal to $|E_1\rangle$ and that lies in the "plane" of Hilbert space picked out by the rays $|E_1\rangle$ and $|E_2\rangle$:



The "reduced" density operator describing the particle is

$$\rho = \langle E_1 | \psi\rangle\langle\psi | E_1\rangle + \langle\sim E_1 | \psi\rangle\langle\psi | \sim E_1\rangle$$

$$= 2^{-1}\langle E_1 | \phi_1\rangle\otimes|E_1\rangle\langle\phi_1|\otimes\langle E_1 | E_1\rangle + 2^{-1}\langle E_1 | \phi_1\rangle\otimes|E_1\rangle\langle\phi_2|\otimes\langle E_2 | E_1\rangle$$
$$+ 2^{-1}\langle E_1 | \phi_2\rangle\otimes|E_2\rangle\langle\phi_1|\otimes\langle E_1 | E_1\rangle + 2^{-1}\langle E_1 | \phi_2\rangle\otimes|E_2\rangle\langle\phi_2|\otimes\langle E_2 | E_1\rangle$$

$$+ 2^{-1}\langle\sim E_1\,|\phi_1\rangle\otimes|E_1\rangle\langle\phi_1|\otimes\langle E_1\,|\sim E_1\rangle + 2^{-1}\langle\sim E_1\,|\phi_1\rangle\otimes|E_1\rangle\langle\phi_2|\otimes\langle E_2\,|\sim E_1\rangle$$

$$+ 2^{-1}\langle\sim E_1\,|\phi_2\rangle\otimes|E_2\rangle\langle\phi_1|\otimes\langle E_1\,|\sim E_1\rangle + 2^{-1}\langle\sim E_1\,|\phi_2\rangle\otimes|E_2\rangle\langle\phi_2|\otimes\langle E_2\,|\sim E_1\rangle$$

$$= 2^{-1}\{\ |\phi_1\rangle\langle\phi_1|\ + \langle E_2\,|E_1\rangle|\phi_1\rangle\langle\phi_2|\ +\ \langle E_1\,|E_2\rangle|\phi_2\rangle\langle\phi_1|\ + |\langle E_2\,|E_1\rangle|^2|\phi_2\rangle\langle\phi_2|$$

$$+ 0\ + 0\ + 0\ + |\langle E_2\,|\sim E_1\rangle|^2|\phi_2\rangle\langle\phi_2|\ \}$$

$$= 2^{-1}\{\ |\phi_1\rangle\langle\phi_1|\ + \langle E_2\,|E_1\rangle|\phi_1\rangle\langle\phi_2|\ +\ \langle E_1\,|E_2\rangle|\phi_2\rangle\langle\phi_1|\ + |\phi_2\rangle\langle\phi_2|\ \},$$

since $|\langle E_2\,|E_1\rangle|^2 + |\langle E_2\,|\sim E_1\rangle|^2 = 1$, as the above diagram shows. (To see this, note that $|\langle E_2\,|E_1\rangle| = \cos\theta$ and $|\langle E_2\,|\sim E_1\rangle| = \sin\theta$.) In summary, the particle's density operator takes the form

$$\rho = c_{11}|\phi_1\rangle\langle\phi_1|\ +\ c_{22}|\phi_2\rangle\langle\phi_2|\ +\ c_{12}|\phi_1\rangle\langle\phi_2|\ +\ c_{21}|\phi_2\rangle\langle\phi_1|,$$

with $|c_{12}|$ and $|c_{21}|$ proportional to $|\langle E_2\,|E_1\rangle|$. In other words, the interference terms (and the resulting interference effects) get more and more "washed out" as the environmental states approach orthogonality. In the "decoherence" limit as $|\langle E_2\,|E_1\rangle| \to 0$, the interference terms disappear entirely, leaving us with the "mixture"

$$\rho_m = c_{11}|\phi_1\rangle\langle\phi_1|\ +\ c_{22}|\phi_2\rangle\langle\phi_2|.$$

In that limit, the particles don't produce an interference pattern on the photographic plate. Instead, they produce a "classical" statistical mixture, exactly *as if* each individual particle

had passed through slit 1 *or* slit 2. The photograph plate would display two dark clumps, one behind slit 1, the other behind slit 2.

But this classical statistical mixture does not automatically imply that each individual particle *really does* pass through one slit or the other, in the classical sense. In fact, we have reasons for denying that classical interpretation. Imagine a double slit experiment in which the experimenter can turn on or turn off the "environment" (perhaps a bunch of air molecules, perhaps a photon bath) between the slits and the photographic plate. Of course, she can't turn off the environment completely; but she can control its "strength." Crucially, let's say the environment is initially "off," and experimenter doesn't decide whether to switch it on until *after* the particle has traversed the slits. If the experimenter leaves the environment "off," then she gets an interference pattern. By the usual interference arguments, this suggests that the particle doesn't pass through slit 1 *or* slit 2, in the classical sense. But now suppose the experiment is repeated; and for each particle, the environment is switched on (after it has traversed the slits). Since this new experiment is exactly the same as the previous one until *after* each particle has already passed through the slits, then either (i) it's still the case that the particle doesn't pass through slit 1 *or* slit 2, in the classical sense; or (ii) each particle somehow "knows" ahead of time whether the experimenter will turn on the environment, and when it knows the environment will be turned on, it decides to pass through slit 1 *or* slit 2 in the classical sense.

Option (ii) implies something far worse than Bell locality violation; it implies backwards in time causation, or else a pre-planned "conspiracy." For this reason, a non-hidden-variable explanation of this experiment should assert that the washing out of the interference pattern results from the particles' interactions with the environment, *not* from the particles' *really* passing through slit 1 or slit 2 in the classical sense.

Of course, in some hidden-variable theories such as Bohm's, the particle does in fact pass through one slit or the other, in the classical sense. The point of my argument is that, if the particle behaves classically at the slits when the environment is turned on, it should also behave classically at the slits when the environment is turned off (assuming the environment isn't "switched on" until after the particle traverses the slits).

I'm dwelling on this experiment in order to make the following interpretive point: Although "decoherence" results in classical statistics (i.e., essentially no interference), we cannot automatically conclude that the underlying individual objects behave classically. Classical behavior implies classical statistics, but the converse fails. The double slit experiment illustrates this point perfectly. Unfortunately, you can find many misunderstandings in the literature that boil down to an unwarranted assumption that classical statistics imply classical behavior (in some sense).

Now that I've explicated decoherence with an example, let me show how it applies to measurement interactions.

Zurek (1993a,b), Joos and Zeh (1985), Bacciagaluppi and Hemmo (1995), and others use general plausibility arguments and worked examples to argue the following: The measuring apparatus undergoes a "dissipative" interaction with its environment. This interaction quickly destroys the coherence between the two branches of the superposition in eq. (1).

(1) **Ideal measurement** $|\phi> = c_1|S_z=+>\otimes|R=+> + c_2|S_z=->\otimes|R=->.$

In this way, the environment picks out the pointer-reading basis.

To see what this means formally, let $|E_+>$ denote the state of the environment (i.e., the rest of the universe) after it interacts with a particle/apparatus system in state

$|S_z{=}+\rangle\otimes|R{=}+\rangle$. Similarly for $|E_-\rangle$. When a particle/apparatus system described by eq. (1) interacts with its environment, the universe evolves into

$$(3) \qquad |\Psi\rangle = c_1|S_z{=}+\rangle\otimes|R{=}+\rangle\otimes|E_+\rangle + c_2|S_z{=}-\rangle\otimes|R{=}-\rangle\otimes|E_-\rangle.$$

As time passes, the environmental states corresponding to different pointer readings quickly approach orthogonality. Formally, as $t\rightarrow\infty$, $\langle E_+|E_-\rangle\rightarrow 0$. For all practical purposes, this "decoherence" takes less than a billionth of a second. In this limit, the reduced density operator describing the particle/apparatus system, found by tracing over the environmental degrees of freedom, is the mixture

$$\rho_m = |c_1|^2|S_z{=}+\rangle\langle S_z{=}+||R{=}+\rangle\langle R{=}+| + |c_2|^2|S_z{=}-\rangle\langle S_z{=}-||R{=}-\rangle\langle R{=}-|.$$

Put roughly, the environment "damps out" the interference terms in the density operator $|\varphi\rangle\langle\varphi|$.

I must stress that according to pure QM, ρ_m describes the particle/apparatus system only because eq. (3) describes the universe, with $\langle E_+|E_-\rangle=0$ in the infinite-time limit. In other words, ρ_m is a "reduced" ("improper") mixture, found by tracing out another subsystem (the environment) with which the system of interest is entangled. As the double slit experiment illustrates, we can't automatically apply a "classical" interpretation to this classical statistical mixture. I'll have more to say about that in the next subsection.

Eq. (1), however, does not describe most real-life (non-ideal) measurements. The more realistic state contains "mistake terms":

(2) **Imperfect measurement**

$$|\phi'\rangle = c_{11}|S_z{=}+\rangle\otimes|R{=}+\rangle + c_{12}|S_z{=}+\rangle\otimes|R{=}-\rangle + c_{21}|S_z{=}-\rangle\otimes|R{=}+\rangle + c_{22}|S_z{=}-\rangle\otimes|R{=}-\rangle,$$

Remember, even if our equipment is flawless, imperfect measurements follow inevitably from wavefunction tails and from environmental fluctuations whose existence is implied by QM. For this reason, a Stern-Gerlach experiment cannot be made ideal, even in principle. Although c_{12} and c_{21} can be made small, they cannot be eliminated.

Why is this important? Because, after the particle/apparatus system interacts with its environment, the final state is given not by eq. (3), but by

$$(4) \qquad |\Psi'\rangle = c_{11}|S_z{=}+\rangle|\otimes|R{=}+\rangle\otimes|E_{++}\rangle + c_{12}|S_z{=}+\rangle\otimes|R{=}-\rangle\otimes|E_{+-}\rangle$$
$$+ c_{21}|S_z{=}-\rangle\otimes|R{=}+\rangle\otimes|E_{-+}\rangle + c_{22}|S_z{=}-\rangle\otimes|R{=}-\rangle\otimes|E_{--}\rangle.$$

As $t\rightarrow\infty$, the environmental states corresponding to different pointer readings approach orthogonality: $\langle E_{++}|E_{+-}\rangle\rightarrow 0$, $\langle E_{-+}|E_{+-}\rangle\rightarrow 0$, $\langle E_{++}|E_{--}\rangle\rightarrow 0$, and $\langle E_{-+}|E_{--}\rangle\rightarrow 0$. But at any finite time, these states are not strictly orthogonal.

Instead of narrowly focusing on how these considerations affect the modal interpretation, let me branch out and explore which other classes of interpretations are helped by decoherence. For now, I'll examine how "decoherence-helped" interpretations interpret eq. (3). Later, we'll see whether these interpretations successfully carry over to non-ideal measurements.

§4.3.2. Decoherence-helped interpretations

Decoherence-helped interpretations--and we'll see that more than one exists--agree on the following:

When the state of the universe takes the form of eq. (3) with the environmental states (nearly) orthogonal, then the pointer reading, or some observable "close" to the pointer reading, is "definite."

Several interpretations fit into this framework, due partly to the different senses in which an observable can be "definite." I'll carve up decoherence-helped interpretations into different classes based upon their answers to two crucial questions:

(A) Does "definite" mean "definite in the absolute, classical sense"?

(B) Does the eigenvector-eigenvalue link hold?

Decoherence interpretation #1: "Definite" = "classically definite," but the eigenvector-eigenvalue link fails.

According to decoherence interpretation #1, an observable (e.g., the pointer reading) can possess a definite value even when the quantum state isn't an eigenstate of the corresponding operator. To pick out which observables become definite, we can rely on the form of the quantum state, as modal interpreters do; or we can invoke formal properties of the relevant interaction Hamiltonian. Although these approaches differ in formal detail, they're both part of the same *program* of letting the interactions between subsystems determine which observables acquire definite values.

Therefore, decoherence interpretation #1 is just a "modal" interpretation, perhaps with a different basis-selection rule.

Decoherence interpretation #2: "Definite" = "classically definite," and the eigenvector-eigenvalue rule holds.

According to this interpretation, which is *prima facie* appealing to many physicists I've spoken with, we can assign an ignorance interpretation to the mixture describing the particle/apparatus system. That is, we can say that the particle/apparatus system *really does* occupy a quantum state corresponding to one of the "legs" of the mixture. But despite its intuitive appeal, decoherence interpretation #2, is inconsistent. Here's why:

By assumption, the eigenvector-eigenvalue rule holds. Therefore, the apparatus has a definite pointer reading only if the quantum state is an eigenstate of **R**. But the quantum state, given by eq. (3) or by eq. (4), is not an eigenstate of **R**.

The inconsistency of decoherence interpretation #2 illustrates D'Espagnat's (1976) point that within pure QM, we cannot assign an ignorance interpretation to an "improper" mixture. It also underscores my conclusion from the double slit experiment, that classical statistics do not imply classical behavior.

Decoherence interpretation #3: Relative-state.

According to this view, the pointer reading becomes definite not in some absolute sense, but relative to its branch of the superposition. Within each branch, the eigenvector-eigenvalue rule holds.

Before discussing whether decoherence solves the ontological problems associated with relative-state and many-world interpretations, I'll briefly discuss what these interpretations are supposed to mean. Both Zurek (1993b) and Zeh (1993), two of the most respected decoherence theorists, stress that their interpretations flesh out Everett's "relative-state" interpretation, not deWitt's many-world interpretation. (See deWitt and Grāham's 1973 anthology.) Although Zurek and Zeh (and Everett) never unambiguously spell out the precise ontology of their interpretations, I'll try to reconstruct an argument that captures (or at least supports) their views.

According to deWitt, after a measurement, each branch of the relevant superposition lives in its own world. If these separate worlds are physically inaccessible to each other, then no interactions can occur between inhabitants of the different worlds, even in principle. Therefore, no "interference" can occur between different branches of the superposition, even in principle. But Zurek and Zeh espouse "pure" QM, according to which Schrödinger's equation governs all state evolution, and hence all interference effects permitted by Schrödinger's equation are possible in principle. For this reason, Zurek and Zeh want the different branches of eq. (3) to inhabit different "realities" that could in principle (though not in practice) interfere. This, along with the radical metaphysics of the many-world view, could partially explain why Zurek and Zeh ally themselves more with Everett than with deWitt.

Unfortunately, the ontology of the Everett-Zurek-Zeh view is unclear. To see why, consider a system in state

$$|\Psi> = c_1|S_z=+>\otimes|R=+>\otimes|E_+> + c_2|S_z=->\otimes|R=->\otimes|E>.$$

We can "see" interference between the two branches of the superposition by measuring $Q = S'\otimes R'\otimes E'$, where **S'** doesn't commute with S_z, **R'** doesn't commute with **R**, and **E'** doesn't commute with $E= a|E_+><E_+| + b|E_-><E_-|$. Although we cannot in practice measure Q, the quantum formalism does not rule out such measurements in principle. Because the "up" and "down" branches can interfere, those branches cannot be said to inhabit "separate" physical realities. Therefore, what it means for an observable to become definite "relative to its branch" is ambiguous. See Albert and Loewer (1988) for a detailed discussion of this objection.

Some physicists downplay the severity of this metaphysical problem. They argue as follows: Sure, when the two branches interfere, it becomes meaningless to assert that the pointer reading is definite relative to its branch. But most of the time, the up and down branches *don't* interfere. During these times, it's unproblematic to claim that the pointer reading is definite, relative to its branch.

This counterargument fails to resolve the ontological ambiguities raised above. When the two branches aren't interfering, do two "copies" everything exist? If not, then in what sense are both measurement results actualized? If so, and if the two branches don't inhabit separate worlds (in deWitt's sense), then how do they co-exist in space and time? In some recent talks, Zurek has taken a more subjectivist stance; minds "live" in one branch or the other, although the world itself doesn't split. But since I have nothing to add to the general arguments for and against relative-state and many-world interpretations, I won't press these questions any further. My point is this: First, the ontology of the relative-state (as opposed to many-world) framework adopted by some decoherence theorists is, at best, ambiguous. Second, decoherence cannot help us to address the metaphysical difficulties facing relative-state and many-world interpretations. If you think these interpretations make no sense, decoherence cannot change your mind.

Summary. In this subsection, I sketched the three most popular decoherence-helped interpretations. (Other such interpretations, though logically possible, have not been developed to my knowledge.) Decoherence cannot help the modal, relative-state, and many-world interpretations fend off general metaphysical criticisms. What decoherence *can* do is help these interpretations pick out a "special" basis. In the modal view, this special basis determines which observables acquire definite values. In the relative-state and many-world view, this special basis determines how physical reality "branches" (in some sense).

I'll now explore to what extent decoherence can help these interpretations select the pointer-reading basis.

§4.3.2. *Decoherence: Selecting the "right" basis*

Any interpretation relying on a "special" basis must specify formal rules that pick out the basis. Since we're trying to explain why measurements result in definite pointer readings, a successful basis-selection rule must choose the pointer-reading basis, or something very "close" to the pointer reading basis, in almost all situations we want to call "measurements." With respect to basis selection, the decoherence-helped interpretations discussed above potentially suffer from two major obstacles: The imperfect measurement problem, and the basis degeneracy problem.

Basis degeneracy problem. This difficulty arises when a basis-selection rule doesn't always choose a unique basis. As an example, consider the usual "modal" rule, also advocated by deWitt for many-world interpretations, of letting the biorthogonal decomposition pick out a special basis. If any two $|c_i|$'s are equal, then the quantum state has multiple biorthogonal decompositions. For instance, consider the particle/apparatus system in state

$$|\varphi\rangle = c_1|S_z{=}+\rangle\otimes|R{=}+\rangle + c_2|S_z{=}-\rangle\otimes|R{=}-\rangle.$$

If $c_1{=}c_2{=}2^{-1/2}$, then $|\varphi\rangle$ can be rewritten as

$$|\varphi\rangle = 2^{-1/2}[|S_x{=}+\rangle\otimes|R'{=}+\rangle + |S_x{=}-\rangle\otimes|R'{=}-\rangle],$$

where

$$|S_X=\pm\rangle = 2^{-1/2}[|S_Z=+\rangle \pm |S_Z=-\rangle]$$

$$|R'=\pm\rangle \equiv 2^{-1/2}[|R=+\rangle \pm |R=-\rangle].$$

Because of this degeneracy, nothing is "special" about the pointer-reading basis, at least, not if we retain the modal basis-selection rule. According to the modal interpretation, if the biorthogonal decomposition isn't unique, then none of the relevant nondegenerate observables acquires a definite value. This is troublesome, at least in principle, because we want S_Z-measurement of a particle initially in state $|S_X=+\rangle$ to yield a definite pointer reading.

Decoherence cannot rescue an interpretation from the basis degeneracy problem. There will always exist coefficients c_1 and c_2 such that a particle initially in state $|\phi\rangle = c_1|S_Z=+\rangle + c_2|S_Z=-\rangle$, after interacting (ideally or non-ideally) with a measuring apparatus that then interacts with the environment, results in a degenerate biorthogonal decomposition with respect to the apparatus. But arguably, the basis degeneracy problem isn't really a problem at all. Of the infinite number of possible initial states of the particle, only a finite number are such that the pointer basis ends up degenerate. And each of those "anomalies" will be "surrounded" in Hilbert space by a continuum of well-behaved initial states. So, no matter how precisely you can prepare your initial states (provided you can't do so with *infinite* precision), we expect such occurrences to happen with zero probability.

Nonetheless, such an occurrence is possible. Should this bother us, i.e., should it count as a "strike" against the modal interpretation (or any interpretation that suffers from basis degeneracy)? David Albert (personal communication) gives us insight into this metaphysical dilemma by raising an analogous example from classical statistical

mechanics. It's possible (though extremely unlikely) for a cold bucket of water to spontaneously boil. This highly counterintuitive--and never experienced--behavior is predicted by the theory. But we don't count it as a strike against the theory, because the theory explains our more commonplace experiences so accurately. Similarly, if quantum mechanics interpreted modally explains our everyday experiences of definite pointer readings, we can't discredit such an interpretation simply because it predicts the possibility of a counterintuitive, never-experienced occurrence, provided it assigns sufficiently low probability to such an occurrence.

A purist could respond that the pointer shouldn't be "allowed" to have an indefinite reading, even in principle. But I tend to side with Albert. The "job" of a physical theory and its interpretation is to explain our experiences. If a theory and interpretation accomplish that goal with simplicity, elegance, breadth, etc., then it's at most a minor aesthetic annoyance if the theory/interpretation predicts the occasional oddball occurrence. If you can forgive classical statistical mechanics that sin, then you should also forgive the modal interpretation of quantum mechanics.

So, the basis degeneracy problem isn't really a problem; or, if it *is* a problem, decoherence can't do anything about it.

Now let's return to the imperfect measurement problem. I'll save until section 4.4 a full discussion of how well the modal interpretation fares. (The suspense builds!) For now, let me take a close at Zurek's basis-selection rule. Instead of relying on the form of the quantum state, he lets the apparatus/environment interaction Hamiltonian, $\mathbf{H}_{int}$, pick out a basis. Here's how:

Let R' denote an arbitrary apparatus observable that doesn't commute with the pointer reading, R. Using "toy" examples, along with general considerations, Zurek argues that $\mathbf{H}_{int}$ commutes with **R**, but does not commute with any **R'**. In rough terms,

the interaction between the apparatus and its environment uniquely picks out the pointer-reading basis. Formally **R** is a "special" pointer-reading observable iff $[\mathbf{H}_{int},\mathbf{R}]=0$.

To see the physical motivation behind this selection rule, pretend that the apparatus's time evolution depends only on its interaction with the environment. In other words, "turn off" the apparatus's internal Hamiltonian, $\mathbf{H}_0$. In this pretend universe, if the apparatus begins in a pointer-reading eigenstate at time t=0, it remains in that eigenstate, because $[\mathbf{H}_{int},\mathbf{R}]=0$. In words, the apparatus/environment interaction leaves the pointer reading undisturbed. By contrast, the environment would knock the apparatus out of an **R**'-eigenstate.

(Because $\mathbf{H}_{int}$ is tremendously complicated in all but the simplest examples, we don't yet know whether Zurek's basis-selection rule avoids the basis degeneracy problem. But based on the examples worked out so far, the prospects look promising.)

Zurek's basis-selection rule cannot suffer from the imperfect measurement problem, because the basis picked out by the apparatus/environment interaction in no way depends on the measurement interaction between the apparatus and the "particle." Formally, the special basis depends only on the apparatus/environment interaction Hamiltonian $\mathbf{H}_{int}$, not on the particle/apparatus interaction Hamiltonian $\mathbf{H}_{measurement}$. Therefore, it doesn't matter how imperfect $\mathbf{H}_{measurement}$ is.

But don't get the idea that Zurek's basis selection rule suffers from no problems. Zurek does not specify when the relative-state "branching" occurs, i.e., at what time the pointer reading acquires a definite value (relative to its branch). Since the environment interacts with the apparatus before, during, and after the measurement, it's not clear when the measurement ends, so to speak. To address this difficulty, Zurek and colleagues must look beyond $\mathbf{H}_{int}$.

Before continuing, it's worth pointing out that, at least in certain crucial idealized cases of decoherence, *the pointer basis picked out by Zurek's rule is precisely the basis asymptotically approached by the modal (biorthogonal) rule.* For instance, in the standard "particle in a harmonic-oscillator heat bath" example, the interaction Hamiltonian is a function of (and therefore commutes with) the particle's position operator. Hence, Zurek's rule selects position as the preferred basis. What about the modal basis-selection rule? The biorthogonal decomposition of the particle-plus-heat-bath picks out a particle operator that asymptotically approaches the position operator. Roughly put, modal interpreters assign a definite value to a physical quantity "very close" to particle position.

At least in some cases, Zurek and the modal interpreters agree about what basis gets selected.[24] The only difference is that Zurek's interpretation picks out the expected pointer-reading basis at all times, whereas the modal interpretation at finite times picks out a basis very close to the pointer-reading basis. So, as mentioned above, Zurek's basis-selection rule solves the imperfect measurement problem. But the modal interpretation still suffers from the problem, unless you think that "close is good enough." Again, I'll discuss the "closeness" question more carefully in section 4.4 below.

When the system of interest is macroscopic, the pointer basis picked out by both interpretations usually corresponds to states of highly-localized position. Or at least, that's the hope.

(Of course, the modal interpretation also picks out a basis *before* decoherence kicks in, and also picks out a basis in situations where decoherence doesn't happen or happens

---

[24]In Elby (1994), I suggested that modal interpreters consider adopting Zurek's basis selection rule.

very slowly. Zurek's has nothing to say about those cases, opening his interpretation up to a charge of "incompleteness.")

We now see that, in many crucial cases, the major difference between Zurek's (relative-state) interpretation and the modal interpretation is metaphysical, not technical. Remember, in the modal interpretation, the "selected" observables take on definite values that are controlled by an independent equation of motion. There's just one "branch" of the universe, and a given definite-valued observable possesses just one of its many possible values. By contrast, in relative state interpretations, all the different possible values of an observable are actualized, in some sense, due to "branching" of the universe.

§4.3.3. Summary

Decoherence cannot help modal, relative-state, or many-world interpretations fend off general metaphysical criticisms. The value of decoherence lies in its ability to pick out a special basis. In the infinite-time limit, modal interpreters and Zurek agree about what "pointer-reading" basis gets picks out. But as I'll discuss below, at finite times, the biorthogonal decomposition picks out a basis close to pointer-reading basis. Furthermore, there exist a nonzero-measure set of initial states such that the biorthogonally-selected basis doesn't get very close to the pointer-reading basis until a noticeable length of time has passed. So, in deciding whether the modal interpretation solves the measurement problem, we must decide whether (i) "Close is good enough," and (ii) It's acceptable that in some cases, nothing even close to the pointer reading takes on a definite value. In the next section, I'll press on exactly these questions.

Section 4.4: Does the modal interpretation, with the help of decoherence, solve the measurement problem?

NOTE TO READERS: A better version of the following argument can be found in a paper by Bacciagaluppi, Elby, and Hemmo, (probably) in the British Journal for the Philosophy of Science, 1996 or 1997.

4.4.1. INTRODUCTION

Taking into account decoherence between systems and their environments, we'll explore how well "modal" interpretations address the measurement problem.

Our argument relies on teasing apart two strands of the measurement problem: the objectification of pointer readings versus the objectification of observers' beliefs about pointer readings. An adequate solution to the measurement problem must explain why a person, after looking at a pointer, perceives its reading as definite. Usually, interpreters of quantum mechanics (QM) assume that if the pointer reading becomes definite, then an observer "automatically" acquires the corresponding definite belief. In modal interpretations, however, the definite values of observables at time t_0 play no role in "choosing" which observables possess definite values at later time t_1. The quantum state alone selects which physical quantities receive definite values. Therefore, the definiteness of a pointer reading does not guarantee that an observer acquires a definite belief about its reading. Whether the person acquires a definite belief depends *entirely* on the biorthogonal decomposition of the overall quantum state in terms of her brain.

For this reason, the debate about whether modal interpretations pick out an observable sufficiently "close" to the pointer reading partially misses the point. Even if

the pointer reading *is* definite, a person might not acquire a definite belief about the pointer reading, in which case the measurement problem remains unsolved.

We show that whether an observer acquires a definite "pointer-reading" belief depends on matters of fact about brain neurophysics. Specifically, we pinpoint necessary conditions that must be satisfied by the physical brain states underlying our definite-belief states, in order for the modal interpretation to assign the observer a definite belief. We then establish the plausibility of these conditions. Finally, we outline what philosophical moves a modal interpreter must make to conclude that the pointer possesses a definite position.

This paper advances two kinds of arguments: Direct arguments about the modal interpretation, and methodological meta-arguments about the kinds of "tests" to which we should subject interpretations of QM. Our direct arguments attempt to show that

•Definite pointer readings do not imply definite beliefs about pointer readings.

•Whether the modal interpretation solves the measurement problem depends on how brains interact with their environment.

•Human brains (and other conceivable conscious beings' brains) probably satisfy two necessary conditions needed for the modal interpretation to work.

•To assign 'definite' pointer positions, modal interpreters must make some nontrivial yet palatable philosophical maneuvers.

Our controversial methodology insists on dragging mental states and physical brain states into the discussion. Despite our lack of knowledge about the relationship between the mental and the physical, a critical evaluation of an interpretation cannot ignore the observer, even when the observer supposedly is not an integral part of the interpretation. We hope to demonstrate the possibility of invoking brain and mental states to make coherent arguments for and against an interpretation.

4.4.2. NOTATION AND PRELIMINARY ASSUMPTIONS

Consider a spin-1/2 particle prepared in a superposition of eigenstates of S_z, the z-component of spin. Denote these eigenstates $|S_z=+\rangle$ and $|S_z=-\rangle$, respectively. The particle interacts with an apparatus designed to measure S_z. Let **R** denote the pointer-reading observable, with eigenstates $|R=+\rangle$ and $|R=-\rangle$, respectively. After the measurement, Diana looks at the pointer. More technically, Diana's brain interacts with the apparatus, largely via the environment of photons. According to QM, the physical state of her brain becomes entangled with the apparatus.

If we want QM to help explain why Diana acquires a definite pointer-reading belief, then we must assume a connection between the mental and the physical. In particular, we assume "supervenience": mental states supervene on physical states.

Some would say our reasoning shouldn't get off the ground, precisely because the relationship between mental and physical is obscure. But many compelling theories of mind assume supervenience. If an interpretation of QM combined with those theories of mind can help to explain our definite beliefs, then the QM interpretation and those theories of mind receive new support. But if a QM interpretation along with supervenience demonstrably cannot explain our post-observation beliefs, then adopting that interpretation practically forces us to renounce supervenience. Unless we're convinced that supervenience fails, not merely skeptical about whether it holds, we shouldn't let an interpretation of QM *make* us renounce supervenience. For these reasons, it's worthwhile to explore whether an interpretation of QM, along with supervenience, can explain why we acquire definite beliefs about pointer readings.

Let |"up",n> denote a brain state corresponding to a "pointer-reading-is-up" belief state. Since many different brain states may fit this description, the index n is needed. Most likely, an uncountable infinity of brain states are "up" states. Our notation won't try to capture this fact, because nothing rides on it.

To define these brain states more carefully, we must consider the "eigenvector-eigenvalue" link, according to which an observable Q possesses a definite value if and only if the system occupies an eigenstate of the corresponding operator **Q**. If the eigenvector-eigenvalue link holds, and if Diana believes the pointer registered up, then by definition her brain occupies state |"up",n> for some n.[25] Similarly for |"down",n>. We do not assume a 1:1 correspondence between physical brain states and belief states. For brevity, we'll often call |"up",n> an "up" belief state, though it's really a physical state underlying the an "up" belief.

Let $|E_{+++n}\rangle$ denote the state of the environment (i.e., the rest of the universe) corresponding to a particle/apparatus/brain in state $|S_z{=}+\rangle\otimes|R{=}+\rangle\otimes|\text{"up"},n\rangle$. To see what this means, suppose the particle initially occupies state $|\Phi\rangle = c_1|S_z{=}+\rangle + c_2|S_z{=}-\rangle$. Suppose the apparatus ideally measures the particle, but Diana "non-ideally" perceives the pointer reading. In other words, when Diana looks at an apparatus in state $|R{=}+\rangle$, she has nonzero probability of perceiving the reading as down; and vice versa. Then the universe ends up in state

$$c_{11}|S_z{=}+\rangle\otimes|R{=}+\rangle\otimes|\text{"up"},1\rangle\otimes|E_{+++1}\rangle + c_{12}|S_z{=}+\rangle\otimes|R{=}+\rangle\otimes|\text{"down"},1\rangle\otimes|E_{++-1}\rangle$$

---

[25]Strictly speaking, |"up",n> might not refer solely to Diana's brain, which constantly exchanges particles with the rest of Diana's body and with its immediate environment. Rather, |"up",n> refers to the state of Diana's brain, body, and perhaps the environment with which she has recently interacted (other than the particle and the apparatus), when Diana believes the pointer registered up.

$$+ c_{21}|S_z=->\otimes|R=->\otimes|\text{"up"},2>\otimes|E_{\_+2}> + c_{22}|S_z=->\otimes|R=->\otimes|\text{"down"},2>\otimes|E_{\_-2}>$$

(1)

State (1) does not represent real life. We use it merely to illustrate notation, and to bolster an argument presented in section 4 below.

4.4.3. TWO MEASUREMENT PROBLEMS

In this section, we tease apart two strands of the measurement problem: The objectification of pointer readings, versus the objectification of belief states.

On most occasions, when Diana looks at the pointer, she perceives its reading as definite, either up or down. But according to the eigenvector-eigenvalue link, a system described by state (1) or any similarly-entangled state does not possess a definite apparatus pointer reading. Similarly, if the eigenvector-eigenvalue link holds, we can't say Diana has a definite "up" belief, because (1) isn't an eigenstate of any operator of the form $\sum_n\sum_m a_{nm}|\text{"up"},n><\text{"up"},m|$.

Some interpretations of QM, such as David Bohm's (see Bohm *et al.* 1987), address the *pointer-objectification problem;* they explain why the pointer reading is indeed definite. For example, according to Bohm, particles always *have* definite positions, and hence pointers have definite centers of mass. Notice that Bohm renounces the eigenvector-eigenvalue link.

Other interpretations attack the measurement problem by showing why our beliefs about pointer readings become definite. According to such interpretations, the pointer reading may be indefinite; but that's acceptable, provided we can explain why people *perceive* pointers as having definite readings. Some versions of Everett's relative-state

interpretation may fit under this umbrella; Albert and Loewer's (1988) many-minds interpretation certainly does. Such interpretations attempt to solve only the *belief-objectification problem,* not the pointer-objectification problem.

Critics of this approach argue as follows: If pointer readings aren't definite, but we nonetheless perceive them to be definite, then nature is fooling us into believing something untrue. In response, the belief-objectificationists argue that we can demand only that an interpretation "save the phenomena" by showing why our beliefs behave as they do.

We need not take a stand on whether explaining belief objectification is sufficient for solving the measurement problem. But it's certainly necessary. To see why, imagine an interpretation according to which a pointer possesses a definite reading, but Diana's brain--incapable of directly "perceiving" the pointer reading--ends up in an "effective" physical state corresponding to a superposition of "up" and "down" belief. Such an interpretation fails, because it cannot explain why Diana believes, with all her heart, that she perceived an up pointer reading. (The fact that Diana sometimes isn't sure does not affect the above argument, because on occasions when Diana *is* sure the pointer registered up, the interpretation must account for her definite belief.)

This point sometimes gets forgotten. Interpreters of QM tend to take it for granted that once the pointer-objectification problem is solved, the belief-objectification problem takes care of itself. As the previous paragraph shows, however, definite pointer readings alone do not entail definite pointer-reading beliefs, and hence do not constitute a full solution to the measurement problem.

In sections 4.4.6 through 4.4.8, we show that modal interpretations teeter near the edge of this trap, but probably don't fall in.

4.4.4 MODAL INTERPRETATION

To the set the stage for section 6, we now briefly review modal interpretations. Whether these interpretations solve the belief-objectification problem depends on matters of fact about brain neurophysics.

Although the modal interpretations of van Fraassen (1979, 1991), Kochen (1985), Dieks (1989, 1994), and Healey (1989, 1995), differ in significant ways, they share enough common elements for us to discuss them as a group. By "modal interpretation," we mean the common elements shared by Kochen, Dieks and Healey.[26]

Modal interpretations retain the linear dynamics of QM, but break the eigenvector-eigenvalue link; an observable can possess a definite value even when the quantum state isn't an eigenstate of the corresponding operator. To pick out which observables receive definite values, the modal interpretation relies on the biorthogonal decomposition theorem. This theorem proves that any state vector describing two subsystems can, for a certain choice of bases, be expanded in the simple "biorthogonal" form $\sum_i c_i|A_i\rangle\otimes|B_i\rangle$, where the $\{|A_i\rangle\}$ and $\{|B_i\rangle\}$ vectors are orthonormal, and are therefore eigenstates of Hermitian operators **A** and **B** associated with subsystems 1 and 2, respectively. Kochen, Healey, and Dieks assert that when $\sum_i c_i|A_i\rangle\otimes|B_i\rangle$ is the unique biorthogonal decomposition of the quantum state with respect to subsystem 1 and 2, then A and B both *have* definite values.[27] (Subsystem 2 can be the "rest of the universe.") So, which

---

[26]van Fraassen's interpretation is developed along somewhat different metaphysical lines.

[27]Some recent formulations of the modal interpretation let the unique spectral resolution of the density operator describing a subsystem pick out the definite observables associated with that subsystem. If the universe occupies a pure state, this basis selection rule is equivalent to the biorthogonal basis selection rule.

observables possess definite values is determined entirely by the *quantum* state. According to Dieks and Healey, a separate set of dynamical laws control the stochastic evolution of the definite values.

Let's apply this "basis-selection" rule to state (1). The pointer reading is definite if that state can be written in the form $|R=+\rangle\otimes|\zeta_1\rangle + |R=-\rangle\otimes|\zeta_2\rangle$, where $\langle\zeta_1|\zeta_2\rangle=0$. By inspection, state (1) already takes that form: $|\zeta_1\rangle\equiv|S_z=+\rangle\otimes(c_{11}|\text{"up"},1\rangle\otimes|E_{+++1}\rangle + c_{12}|\text{"down"},1\rangle\otimes|E_{++-1}\rangle)$ is orthogonal to $|\zeta_2\rangle\equiv|S_z=-\rangle\otimes(c_{21}|\text{"up"},2\rangle\otimes|E_{--+2}\rangle + c_{22}|\text{"down"},2\rangle\otimes|E_{--2}\rangle)$, because $\langle S_z=+|S_z=-\rangle=0$. So, if measurements are ideal, then the modal interpretation solves the pointer-objectification problem.

Does this mean the belief-objectification problem is also solved? Not necessarily. Intuitively, we want to say that when Diana looks at the pointer, she directly perceives the pointer's reading (when it's definite). But according to modal interpretations, that doesn't necessarily happen. When subsystems 1 and 2 interact, the definite values associated with subsystem 1 do not determine which observables associated with subsystem 2 become definite, or vice versa. The definite value associated with the pointer does not directly "cause" Diana to acquire a definite belief about the pointer. Whether Diana acquires a definite belief depends *entirely* on the biorthogonal decomposition of the resulting quantum state with respect to her brain. Of course, *if* Diana forms a definite belief, *then* the dynamics of the definite values can ensure with high probability that her belief mirrors the pointer's actual reading.

In state (1), the observer possesses a definite "up" or "down" belief if the four environmental states are precisely orthogonal. But those states won't be precisely orthogonal. Indeed, if |"up",1> and |"down",1> are not macroscopically distinct--for instance, if memories are stored in atomic spins--then $|E_{+++1}\rangle$ and $|E_{++-1}\rangle$ won't be

even nearly orthogonal. In that case, the biorthogonal decomposition of state (1) with respect to brain states almost certainly picks out states not even close to "up" and "down" states, unless c_{12} and c_{21} are tiny. We're not claiming that memories are stored in spins and that perceptions are highly imperfect. In fact, we argue below that because brains satisfy certain (contingent) conditions, "up" and "down" states *do* get selected in almost all real-life observations. Nonetheless, our current point is this: In the modal interpretation, the definiteness of the pointer reading does not entail the definiteness of Diana's belief about the pointer reading.[28]

Let us summarize the main points of this section.

Point 1: Whether the modal interpretation explains belief objectification depends on matters of fact about brain neurophysics, i.e., on how brains interact with pointers and with the environment. This conclusion continues to hold when we consider a realistic particle/apparatus/brain/environment state.

Point 2: Point 1 holds even though state (1) assigns the pointer a definite reading. Therefore, it's possible for the modal interpretation to solve the pointer-objectification problem without solving the belief-objectification problem. The modal interpretation illustrates our general point that definite pointer readings alone do not entail definite pointer-reading beliefs, and hence do not constitute a full solution to the measurement problem. (Notably, this conclusion applies also to Bub's (1992) modal theory, which ascribes *a priori* definite readings to pointers.)

For this reason, the ongoing debate about whether modal interpretations assign a definite value to an observable "sufficiently close" to the pointer reading misses a crucial

---

[28]This conclusion, you can quickly convince yourself, applies equally well to essentially all interpretations that attempt to solve both the belief-objectification and the pointer-objectification problems.

point By parrying the "imperfect measurement" challenge posed by Albert and Loewer (1990) and by Elby (1993), modal interpreters can establish *only* that the modal interpretation adequately addresses the pointer-objectification problem. And that's not good enough.

4.4.5. IMPERFECT MEASUREMENTS AND PERCEPTIONS

So far, we've established that the modal interpretation might explain pointer objectification without also explaining belief objectification. To pursue these issues, we must write down the actual (non-idealized) post-measurement, post-perception quantum state of the particle/apparatus/person/environment.

First, we'll briefly reiterate Elby's (1993) argument that wavefunction tails and inevitable environmental "fluctuations" prevent the pointer-reading from becoming perfectly correlated with the particle's z-component of spin. Then, we'll argue that Diana's perceptions are imperfect, for the same reasons. As a result, her "up" and "down" belief states do not become perfectly correlated with up and down pointer readings. Finally, taking into account these imperfections, we'll write down the overall quantum state. It's ugly.

Imperfect measurements. QM implies that measurements of some observables are non-ideal, no matter how well we design our equipment. To see why, imagine a standard Stern-Gerlach experiment; a spin-1/2 particle passing between two magnets gets deflected up or down (roughly speaking). Two separate "photographic" plates, one in the up path, the other in the down path, await the particle. The particle hits one of the two plates and produces a dot.

According to Schrödinger's equation, even if the particle was initially localized within a bounded volume, its wavefunction immediately "spreads out" so as to cover all space. (Even in relativistic QM, the wavefunction spreads over the whole forward light cone.) Therefore, an initially spin-up particle has non-zero probability of hitting the "down" plate, because the tail of the up-deflected wavefunction reaches the down plate.

Imperfect measurements also result from environmental fluctuations, according to QM. Suppose the particle "reaches" the up plate. It might produce no dot, because it has a nonzero probability of tunneling through the plate or embedding itself without producing a dot. Also, stray particles hitting the down plate have nonzero probability of producing a dot. Of course, these "fluctuation" probabilities are ridiculously small. Nonetheless, environmental fluctuations guarantee that upon measuring a spin-up particle, we may end up with a dot only on the down plate.

In brief, due to wavefunction tails and environmental fluctuations, the post-measurement state of the particle/apparatus/environment system is

$$c_{11}|S_z=+\rangle\otimes|R=+\rangle\otimes|E_{++}\rangle + c_{12}|S_z=+\rangle\otimes|R=-\rangle\otimes|E_{+-}\rangle$$
$$+ c_{21}|S_z=-\rangle\otimes|R=+\rangle\otimes|E_{-+}\rangle + c_{22}|S_z=-\rangle\otimes|R=-\rangle\otimes|E_{--}\rangle, \qquad (2)$$

where c_{12} and c_{21} are small but nonzero. Actually, c_{12} and c_{21} can be large for poorly designed or broken measuring devices. The environmental states corresponding to different pointer readings are very nearly, *but not exactly,* orthogonal.

Imperfect perceptions. What happens when Diana looks at the pointer? Her eyes interact with photons, some of which previously interacted with the apparatus. According to QM, this (mediated) interaction yields imperfect correlations between the pointer reading and Diana's perception thereof, for the reasons just discussed.

For instance, suppose Diana observes a pointer in state |R=+>. The photons streaming into her eyes from the pointer have small but nonzero probability of failing to "activate" the appropriate receptors on her retina. Similarly, thermal fluctuations could cause certain nerve cells to fire so as to "simulate" seeing a down pointer. Again, QM assigns an infinitesimal but nonzero probability to this eventuality. So, Diana has nonzero probability of acquiring a "down" belief state after observing an up pointer. Her perceptions are imperfect.

Now actually, when Diana "misperceives" an up pointer, she might not acquire a "down" belief. She might acquire no belief at all, or might end up in a superposition of these possibilities. To capture these options, let |~"up",n> denote a brain state that isn't an "up" state. Keep in mind that |~"up",n> does not always correspond to believing the pointer is "not up." It corresponds to not believing the pointer is "up." So, the "down" states are only a subset of the |~"up",n> states.

Of course, so-called "psychological" factors may contribute far more to Diana's imperfect perceptions. A critic could say that, if such psychological factors exist, then Diana's brain does not occupy a proper "ready state" to "measure" the pointer reading. We've just shown that according to QM, no matter how "ready" Diana is to perceive accurately, she will sometimes err. Nonetheless, our everyday experiences strongly suggest that perceptions work well over a broad range of "initial" brain states. Diana's brain need not occupy one particular ready state to perceive the pointer with high accuracy. Almost any of the brain states corresponding to "paying close attention to the pointer" will do.

Because of perceptual imperfection, when Diana looks at the particle/apparatus/environment described by state (2), the system evolves into

$$c_{111}|S_z{=}{+}\rangle\otimes|R{=}{+}\rangle\otimes|\text{"up"},1\rangle\otimes|E_{+++1}\rangle + c_{112}|S_z{=}{+}\rangle\otimes|R{=}{+}\rangle\otimes|\sim\text{"up"},1\rangle\otimes|E_{++-1}\rangle$$
$$+ c_{121}|S_z{=}{+}\rangle\otimes|R{=}{-}\rangle\otimes|\sim\text{"down"},2\rangle\otimes|E_{+-+2}\rangle + c_{122}|S_z{=}{+}\rangle\otimes|R{=}{-}\rangle\otimes|\text{"down"},2\rangle\otimes|E_{+--2}\rangle$$
$$+ c_{211}|S_z{=}{-}\rangle\otimes|R{=}{+}\rangle\otimes|\text{"up"},3\rangle\otimes|E_{-++3}\rangle + c_{212}|S_z{=}{-}\rangle\otimes|R{=}{+}\rangle\otimes|\sim\text{"up"},3\rangle\otimes|E_{-+-3}\rangle$$
$$+ c_{221}|S_z{=}{-}\rangle\otimes|R{=}{-}\rangle\otimes|\sim\text{"down"},4\rangle\otimes|E_{-+4}\rangle + c_{222}|S_z{=}{-}\rangle\otimes|R{=}{-}\rangle\otimes|\text{"down"},4\rangle\otimes|E_{--4}\rangle. \quad (3)$$

The coefficients corresponding to "misperceptions," such as c_{112} and c_{121}, are extremely small. The coefficients corresponding to "mismeasurements" but not misperceptions, such as c_{122} and c_{211}, could be small or large, depending on the sloppiness of the measurement. See Elby (1994) and especially Bacciagaluppi and Hemmo (1994, 1995) for some of the technical details leading to state (3). Crucially, two "different" brain states are very close if they correspond to particle/apparatus states that differ only in the spin of the particle. For instance, consider the c_{111} and c_{211} terms. Because some of the photons that interact with the spin-1/2 particle eventually reach the observer, the observer's brain state "depends" on the particle's spin. But this dependence is negligible. Formally, $\langle\text{"up"},1|\text{"up"},3\rangle\approx 1$.

At this point, we've written the non-idealized, post-measurement, post-perception state of the whole system. Now we can use this state to explore whether the modal interpretation solves the belief-objectification problem and the pointer-objectification problem.

4.4.6. ARE BELIEF STATES DEFINITE?

In this section, we'll argue as strongly as possible that state (3) does not assign Diana a definite "up" or "down" belief state, according to the modal interpretation. (In

sections 7 and 8. we'll argue the other side.) Before advancing these arguments, however, we must explore some characteristics of belief states in the context of the orthodox interpretation.

Assume the eigenvector-eigenvalue rule holds. We've defined the |"up",i> states as follows: If Diana believes that the pointer reading is up, then the quantum mechanical state of her brain is |"up",n> for some n.

Suppose Diana's brain occupies a superposition of "up" states, $\sum_i g_i$|"up",i>. Does she believe the pointer registered up? Not necessarily. A superposition of "up" states necessarily equals another "up" state just in case a Hermitian operator corresponds to "up" belief, i.e., just in case a projection operator $\mathbf{P}_{up}$ exists with the property $\mathbf{P}_{up}$|"up",i>=|"up",i> for all i.

Now suppose Diana's brain occupies state g_1|"up",1>+h_1|~"up",1>, where h_1 is tiny. In words, Diana's brain occupies a state very close to a definite "up" state. Does Diana believe the pointer registered up? If she does, then her brain occupies state |"up",n> for some n≠1, *because by definition, if she believes the pointer reading was up, then she occupies an "up" brain state.* That is, Diana perceives the pointer as "up" only if g_1|"up",1>+h_1|"down",1>=|"up",n> for some n. In other words, when we say g_1|"up",1>+h_1"down",1> is sufficiently close to a definite "up" belief state, we're really saying that g_1"up",1>+h_1|"down",1> *is* a definite "up" belief state. Put another way, if Diana can't distinguish between her beliefs when her brain occupies |"up",1> versus when her brain occupies g_1|"up",1>+h_1|"down",1>, then g_1|"up",1>+h_1|"down",1> is just as definite an "up" belief state as |"up",1> is, and hence g_1|"up",1>+h_1|"down",1>=|"up",n> for some n≠1.

Now for the punch line. Suppose Diana's brain occupies state $\sum_i g_i$|"up",i>+$\sum_i h_i$|~"up",i>, where all the h_i and most of the g_i coefficients are tiny

(though not strictly 0). Does Diana believe the pointer registered up? By the argument of last paragraph, only if $\sum_i g_i|\text{"up"},i\rangle+\sum_i h_i|\sim\text{"up"},i\rangle = |\text{"up"},n\rangle$ for some n. To explore whether this equality holds, we must consider two cases.

Case 1: No $\mathbf{P}_{up}$ operator exists. Therefore, the superposition $\sum_i g_i|\text{"up"},i\rangle$ is not necessarily an "up" state. Therefore, $\sum_i g_i|\text{"up"},i\rangle+\sum_i h_i|\sim\text{"up"},i\rangle$ is not necessarily an "up" state, even if the h_i's vanish entirely.

Case 2: A $\mathbf{P}_{up}$ operator exists, and hence the "up" states live in a closed subspace. But then, $\sum_i g_i|\text{"up"},i\rangle+\sum_i h_i|\sim\text{"up"},i\rangle$ does *not* inhabit the "up" subspace, no matter how small $\sum_i|h_i|^2$ is, provided it's nonzero. Therefore, $\sum_i g_i|\text{"up"},i\rangle+\sum_i h_i|\sim\text{"up"},i\rangle$ does not equal $|\text{"up"},n\rangle$ for any n; Diana does not believe the pointer registered up.

We've just shown that whether or not a Hermitian operator corresponds to "up" belief, $\sum_i g_i|\text{"up"},i\rangle+\sum_i h_i|\sim\text{"up"},i\rangle$ is not necessarily a definite "up" belief state, no matter how small $\sum_i|h_i|^2$ is.

Why is this relevant to the modal interpretation? When we biorthogonally decompose state (3) with respect to Diana's brain, the "selected" observables have eigenstates of the form $\sum_i g_i|\text{"up"},i\rangle+\sum_i h_i|\sim\text{"up"},i\rangle$ and $\sum_i g_i|\text{"down"},i\rangle+\sum_i h_i|\sim\text{"down"},i\rangle$, where $|g_i|\approx 1$ for one i, and all the other coefficients are small (though nonzero).[29] As just shown, this state of affairs does *not* necessarily correspond to Diana's having a definite "up" belief or "down" belief (or "unsure" belief). Therefore, the modal interpretation might not solve the belief-objectification problem.

4.4.7. RESCUING DEFINITE BELIEFS: CLOSENESS

[29]States of the form $\sum_i g_i|\text{"?"},i\rangle+\sum_i h_i|\sim\text{"?"},i\rangle$ also get picked out, where $|\text{"?"},i\rangle$ corresponds to a mental state of being unsure. This doesn't affect our argument.

But all is not lost. Section 6 focuses us on some characteristics that brains must possess in order for the biorthogonal decomposition to pick out "up" and "down" (and "unsure") beliefs. In this section, we discuss why the biorthogonally selected brain states are extremely close to definite "up" and "down" belief states. Then, in section 8, we show why "close" is probably good enough to explain our definite beliefs.

As just noted, when state (3) is decomposed, "effective" brain states of the form

$$\Sigma_i g_i|\text{"up"},i\rangle+\Sigma_i h_i|\sim\text{"up"},i\rangle \qquad (*)$$

or the analogous "quasi-down" states get picked out, where (say) $|g_1|$ is close to one, and all other coefficients are close to zero. By "effective" state, we mean the state that corresponds to a definite possessed property, according to the interpretation under consideration. The modal interpretation uses biorthogonal decompositions to choose effective states. The more closely $|g_1|$ approaches one, the "closer" state (*) is to a definite "up" state, namely $|\text{"up"},1\rangle$. How closely $|g_1|$ approaches one depends on two factors:

(a) How well brain states become correlated with the corresponding pointer-reading states; and

(b) To what extent the brain's "environment," such as thermal degrees of freedom, brings about decoherence between the various "up" and "down" belief states superposed in state (3).[30]

---

[30]The "very close" brain states, such as $|\text{"up"},1\rangle$ and $|\text{"up"},3\rangle$, need not decohere for factor (b) to "work." Which is good, because "close" states can't decohere.

At first glance, factor (a) might seem to downplay the role of the environment. For suppose that (nearly) orthogonal brain states become extremely well correlated with orthogonal pointer-reading states. In other words, suppose our perceptions are excellent. Then, even if the environmental states in (3) were highly non-orthogonal, the relevant biorthogonal decompositions would pick out apparatus states extremely close to pointer-reading eigenstates, and brain states of the form (*) with $|g_1|$ extremely close to one. So, if observers have excellent perceptions, then environmental decoherence seems to play no role in ensuring that the apparatus and brain effective states are extremely close to the ones we want. This conclusion violates an emerging orthodoxy about the importance of the environment.

A closer examination, however, reveals the orthodoxy not to be threatened by factor (a). Recall that the environment mediates the interaction between the pointer and Diana's brain. To see why that mediated interaction "needs" decoherence, let |up photons> and |down photons> denote the state of the photons reaching Diana's eyes from an up pointer and down pointer, respectively. Suppose that |up photons> becomes extremely well correlated with "up" belief. Then, since the relevant interaction Hamiltonians are linear, it follows that the "misperception" terms

$$|...\rangle \otimes |R=\rightarrow\rangle \otimes |\text{"up"},i\rangle \otimes |...\rangle$$

have total "probability"

$$|\langle \text{up photons} \mid \text{down photons} \rangle|^2.$$

In other words, the squares of the coefficients of the |...>⊗|R=->⊗|"up",i>⊗|...> terms add up to this "probability" value.[31] Therefore, if the up and down photon states aren't nearly orthogonal, then Diana's beliefs become poorly correlated with the pointer reading.

This argument proves that beliefs become well correlated with pointer readings only to the extent that the corresponding environmental states are nearly orthogonal--that is, only to the extent that the environment decoheres the pointer-reading eigenstates. Indeed, decoherence must "work" so well that near-orthogonality applies even to small spatial regions of the environment. So, decoherence plays a key role in guaranteeing closeness.

Of course, |up photons> doesn't always lead to "up" beliefs, for the quantum mechanical reasons discussed above. But because our eyes receive billions of photons from pointers, not all of those photons need to be "perceived" properly in order for our eyes to form the right image. Neurons are "well-designed" and macroscopic. The odds are infinitesimal that enough neurons will misfire so as to misread the pointer. In brief, environmental decoherence, when combined with the macroscopic, redundant nature of our visual systems, ensures that perception states (and presumably, the corresponding belief states[32]) become extremely well correlated with pointer-reading states.

For this reason, factor (a) is all that's needed to ensure that Diana acquires a definite belief. Factor (b), by contrast, is needed to ensure that definite belief states are "stable." For suppose that brain/environment decoherence picked out a new basis not close to "up" and "down" states. Then, soon after Diana observes the pointer, the biorthogonal

---

[31]In state (3), the |...>⊗|R=->⊗|"up",i>⊗|...> terms are "built into" the |...>⊗|R=->⊗|~"down",i>⊗|...> terms.

[32]We must assume that when the visual cortex forms a representation of an up pointer, and Diana is "paying attention" to this image, then she almost always acquires an "up" belief.

decomposition of the particle/apparatus/brain/environment with respect to Diana's brain would select those new states, not the definite belief states. In other words, decoherence would "knock" her brain from an "up" or "down" effective belief state into a new effective state. But empirically, we know that beliefs persist far longer than the "decoherence time." For this reason, the modal interpretation can solve the belief-objectification problem only if the Belief Stability Condition holds:

Belief Stability Condition: Brain/environment decoherence picks out (states very close to) definite belief states.

This condition, we must emphasize, does not follow from the modal interpretation. Rather, it's an empirical hypothesis about brains, the falsehood of which would doom the modal interpretation. We'll briefly discuss two reasons for affirming this condition.

First, even though neuroscientists don't understand the details of memory formation and storage, they strongly suspect that the spatial distribution of chemical compounds plays a key role. Roughly put, memories are probably "stored" not in the spins or energy states of molecules, but in their positions. If "up" versus "down" memories indeed correspond to billions of molecules occupying even slightly different positions, then thermal-bath decoherence alone would probably ensure the stability of those memories. Of course, no has proven that thermal-bath decoherence in a complicated potential picks out a basis close to the position basis; but this assumption seems reasonable, given the decoherence results to date. In brief, preliminary data from neuroscience and from decoherence theory suggest (but do not prove) that the Belief Stability Condition holds.

Second, if this condition fails, then it's hard to see how *any* interpretation of QM--except explicitly dualistic ones--can account for the stability of our beliefs. You can confirm that if the Belief Stability Condition fails, then Bohm's theory, relative-state interpretations, and collapse models run into trouble. If this condition fails, then brain/environment decoherence would not prevent quantum interference between different belief states, precisely the kind of interference we never seem to "experience."

In this section, we argued that except in rare pathological cases of near-degeneracy, the modal interpretation picks out brain states of the form (*) that are extremely close to definite "up" and "down" states.[33] Pointer/environment decoherence ensures that the photons reaching Diana's eyes from an up pointer are sufficiently "orthogonal" to photons reaching her eyes from a down pointer that her perception becomes well correlated with the pointer reading. But brain/environment decoherence entails that those beliefs persist only if the Belief Stability Condition holds. As just discussed, we have independent reasons for thinking it does.

4.4.8. IS 'CLOSE' GOOD ENOUGH FOR BELIEFS?

For easy reference, we'll rewrite the effective state (*) picked out by the modal interpretation:

$$\Sigma_i g_i|\text{"up"},i\rangle+\Sigma_i h_i|\sim\text{"up"},i\rangle, \qquad (*)$$

---

[33]If we consider measurements of continuous variables, then things get more complicated. Then the g's become a probability density, g(i), with i a continuous parameter. This probability density is sharply peaked. In other words, the only "up" states appearing with non-negligible probability density in (*) are all very close to each other. (Recall that two states are close if their inner product is nearly 1.)

where (say) $|g_1|\approx 1$, and all other coefficients are close to zero. As emphasized in section 6, this state corresponds to a definite "up" belief only if (*)= |"up",n> for some $n\neq 1$. Now, we want Diana to acquire a definite belief no matter what pre-measurement state the particle occupies. Therefore, since the particle's initial state partly "controls" the g_i's and h_i's in (*), that state had better be a definite "up" state over a broad range of g_i's and h_i's such that one $|g_i|$ is close to one. That is, the modal interpretation can solve the belief-objectification problem only if the Belief Imperturbability Condition holds:

Belief Imperturbability Condition: "Many" of the states very close to |"up",i> must themselves be "up" states.

Like the Belief Stability Condition, the Belief Imperturbability Condition does not follow from general formal considerations or from the modal interpretation. It's another empirical hypothesis that must be satisfied, or else the modal interpretation cannot explain belief objectification.

Some philosophers would argue that the adequacy of a solution to the measurement problem should not depend so crucially on contingent facts about brain architecture. We agree that if an interpretation depends on *delicate* and peculiar facts about human neurophysiology, then we have grounds for complaint. But if the Imperturbability Condition holds for any conscious being likely to evolve (or get built), then smart lizards, artificially-intelligent computers, and Martians could all agree that the modal interpretation fares well.

Our everyday experiences give us ample reason to affirm this condition. For suppose Diana looks at the pointer reading and then accidentally bumps her head. If she bumps it hard enough, she will forget what she observed. So, a head-bump disturbs the

"aspect" of Diana's effective brain state on which definite pointer-reading beliefs supervene. (Although cognitive science has no idea what "aspect" means, our argument applies to most conceivable senses of "aspect," including "degrees of freedom") Now suppose Diana bumps her head softly. Presumably, she thereby disturbs the relevant "aspect" of her effective brain state, though less severely. More formally, the effective brain state picked out by the biorthogonal decomposition presumably gets knocked into a nearby state. But after soft head-bumps, Diana almost always retains (remembers) her old pointer-reading belief. Similar arguments apply to most other conscious beings, because they too could not avoid occasional bumps. These considerations strongly suggest that the Belief Imperturbability Condition holds. By the way, many other interpretations of QM may rest on this condition, too.

At first glance, the Belief Imperturbability Condition seems dispensable for modal interpreters, provided the Belief Stability Condition holds. Here's the reasoning, put roughly: Even if a head-bump knocks Diana out of an "up" effective brain state, decoherence quickly knocks her back into a definite-belief effective brain state, before she notices the disturbance. And the dynamics of the possessed values can ensure that, with high probability, she ends up with the same belief she held initially. To see the flaw in this argument, let |pre> and |post> denote the effective brain state picked out by the biorthogonal decomposition and by the dynamics of the possessed values immediately before and immediately after the head-bump. The above argument implicitly assumes that only |pre>, but not |post>, is part of the basis picked out by decoherence. We have no reason to assume this. Perhaps |pre> and |post> are degenerate basis vectors selected by decoherence.

To summarize: We've now spelled out the Belief Stability and Belief Imperturbability Conditions. Although they seem similar or redundant, these conditions

are logically independent. If either of them fails, the modal interpretation can't solve the belief objectification problem.[34] But we have independent reasons to affirm both these conditions. In this sense, the modal interpretation passes a crucial "test" regarding its ability to solve the measurement problem.

4.4.9. CONCLUSION

This paper teased apart two strands of the measurement problem, the pointer-objectification problem and the belief-objectification problem. Many interpreters of QM take it for granted that a solution to the pointer-objectification problem automatically addresses the belief-objectification problem as well. But this isn't true in the modal interpretation, because a brain does not directly perceive the actual value of the pointer reading. Whether the observer acquires a definite belief depends entirely on the biorthogonal decomposition of the state resulting from the quantum mechanical interaction between the brain, pointer, and environment. *If* the quantum state "gives" the observer a definite belief, *then* the dynamics of the possessed values can ensure that she almost certainly acquires the "correct" belief.

Because measuring devices don't ideally measure particles, and brains don't ideally "measure" pointer readings, the final quantum state of the particle/apparatus/brain/environment system is a mess, state (3). The modal interpretation, applied to state (3), assigns a definite value to an observable whose eigenstates take the form $\sum_i g_i|\text{"up"},i\rangle+\sum_i h_i|\sim\text{"up"},i\rangle$ (or the "down" analog), where $|g_i|\approx 1$ for one i and all other coefficients are nearly zero. This state of affairs corresponds

---

[34]Actually, as just implied, Belief Imperturbability could fail provided most accessible $|post\rangle$ states are not part of the decoherence basis.

to a definite "up" or "down" belief state only if most brain states very close to an "up" state are themselves "up" states. But the persistence of our beliefs in the face of brain-jostling strongly suggests that this imperturbability condition holds. Furthermore, by arguing that an object's macroscopic "position" supervenes on an object's physical state but does not correspond 1:1 to the quantum mechanical position operator, we can make sense of the claim that the pointer has a definite position when an observable "close" to the pointer reading is definite. Overall, the modal interpretation appears to fare well with respect to both pointer and belief objectification.

Section 4.5: Holism in the modal interpretation

In my previous chapters, I've used no-go results to argue that nature incorporates a kind of holism. But I've been sketchy about what "holism" *is* and what philosophical work it does. In this section, I'll show in exactly what sense "holism" gets incorporated into the modal interpretation. Since the modal interpretation is viable, and since other viable interpretations (such as Bohm's) also incorporate a kind of holism, this discussion could shed some (perhaps veiled) light on how holism is a feature of the world.

My philosophical points will closely follow those of Healey (1994). But my discussion will clear up some technical oversimplifications that plague Healey's presentation.

§4.5.1. Singlet-state: Before measurement

Continue to consider two spin-1/2 particles in their singlet state. Particle 1 passes through a z-aligned Stern-Gerlach magnet and gets deflected up or down. It eventually hits a "photographic" plate in the "up" or "down" path. All of this happens *before* particle 2 gets measured.

If particle 1 is measured to have spin up, particle 2 now has (conditional) probability unity of yielding spin down. Therefore, according to EPR's necessary condition for an "element of reality," there exists an element of reality corresponding to the definite z-spin of particle 2. If quantum mechanics is "complete," that element of reality didn't exist *until* particle 1 was measured. Therefore, according to EPR, an element of reality associated with particle 2 was *nonlocally* created by measuring particle 1. And this nonlocality is metaphysically unacceptable.

The modal interpretation addresses this "paradox" as follows. As we'll see, an element of reality corresponding to the z-spin of particle 2 is indeed created when particle 1 gets measured. But that element of reality is not a property of particle 2. Rather, it's a holistic property of the two-particle system. Measuring particle 1 doesn't create an element of reality associated with a *separate,* distant system. Rather, it creates an element of reality associated with an "extended" system of which particle 1 is a "part." The nonlocal connection between particles 1 and 2 is a holistic nonseparability, *not* a superluminal signal or other classically "causal" agent.[35]

To see how this holistic interpretation of Bell-nonlocality follows from the modal interpretation, first consider the particles before particle 1 reaches the magnet. Their state is $|\text{singlet}\rangle = 2^{-1/2}\{|S_z{=}+\rangle_1 \otimes |S_z{=}-\rangle_2 - |S_z{=}-\rangle_1 \otimes |S_z{=}+\rangle_2\}$. Since this biorthogonal decomposition isn't unique, no nondegenerate observable associated with either particle takes on a definite value. In other words, neither particle *has* a definite spin component in any direction. But the two-particle system as a whole possesses a definite "spin-correlation" property $P_{|\text{singlet}\rangle}$ corresponding to the Hermitian operator $\mathbf{P}_{|\text{singlet}\rangle} = |\text{singlet}\rangle\langle\text{singlet}|$. This property encodes the perfect spin anticorrelations between the two particles. And crucially, this spin-anticorrelation property $P_{|\text{singlet}\rangle}$ doesn't "pick out" any direction. The particles' spins are anticorrelated not only in the z-direction, but also in the x-direction, y-direction, and so on.

I can't overemphasize the fact that $P_{|\text{singlet}\rangle}$ is a holistic property, by which I mean it can't be reduced to (or "built up from") the properties of the individual particles. In philosopher's lingo, the property of the whole does not supervene on the properties of its parts. This departs radically from the reductionistic metaphysics of classical physics. In a Newtonian universe, a two-particle system has zero net angular momentum *because*

---

[35]I'll address causation more carefully in chapter 5.

the two individual particles have no angular momentum, or *because* the particles have equal and opposite angular momenta.

Now consider the particles after particle 1 has passed through the Stern-Gerlach magnet, but before it reaches a photographic plate. This is the "intermediate" stage of the measurement process. Let $|\phi_{up}\rangle$ and $|\phi_{down}\rangle$ denote the spatial wavefunction of particle 1 when it gets deflected up and down, respectively. The "intermediate" state of the system is

$$|\text{intermediate}\rangle = 2^{-1/2}\{|S_z=+\rangle_1 \otimes |S_z=-\rangle_2 \otimes |\phi_{up}\rangle - |S_z=-\rangle_1 \otimes |S_z=+\rangle_2 \otimes |\phi_{down}\rangle\}.$$

To find the definite properties associated with each subsystem, we must look at the relevant biorthogonal decompositions. Well, with respect to the spin of particle 1, |intermediate> is biorthogonally decomposed. (That's because $|S_z=-\rangle_2 \otimes |\phi_{up}\rangle$ is orthogonal to $|S_z=+\rangle_2 \otimes |\phi_{down}\rangle$, since opposite spin states are orthogonal.) But the decomposition isn't unique, because the expansion coefficients are degenerate: $c_1=c_2= 2^{-1/2}$. So, particle 1 still doesn't have a definite spin component in any direction. By similar reasoning, the same goes for particle 2.

But things get more interesting when we take as our "subsystem of interest" the spin characteristics of the two-particle system as a whole. |intermediate> is *not* biorthogonally decomposed with respect to that subsystem. Here's why: A biorthogonal decomposition takes the form $\sum_i c_i |A_i\rangle \otimes |B_i\rangle$, where $|A_i\rangle$ are orthogonal states describing the subsystem of interest, and $|B_i\rangle$ are orthogonal states describing everything else. Since the subsystem of interest is the spin-characteristics of the two-particle system, the $|A_i\rangle$ states are $|S_z=+\rangle_1 \otimes |S_z=-\rangle_2$ and $|S_z=-\rangle_1 \otimes |S_z=+\rangle_2$, which are indeed orthogonal. But the $|B_i\rangle$ states, $|\phi_{up}\rangle$ and $|\phi_{down}\rangle$, are *not* orthogonal, because of

their overlapping wavefunction tails, as discussed in section 4.2 above. So, |intermediate> is *not* biorthogonally decomposed with respect to the subsystem of interest. To find the definite-valued observable associated with this subsystem, we must re-write |intermediate> in terms of a new, biorthogonal basis. (By the biorthogonal decomposition theorem, such a basis exists.) Since $|\phi_{up}\rangle$ and $|\phi_{down}\rangle$ are *nearly* orthogonal, the biorthogonal decomposition takes this form:

$$|\text{intermediate}\rangle = d_1|A_1\rangle\otimes|\phi'_{up}\rangle + d_2|A_2\rangle\otimes|\phi'_{down}\rangle.^{36}$$

Since we're "close to a degeneracy"--i.e., since $|d_1|$ and $|d_2|$ are almost equal, or perhaps exactly equal--$|A_1\rangle$ will *not* in general be close to $|S_z{=}{+}\rangle_1\otimes|S_z{=}{-}\rangle_2$. However, $|A_1\rangle$ and $|A_2\rangle$will *always* take the form

$$|A_1\rangle = \cos\theta\ |S_z{=}{+}\rangle_1\otimes|S_z{=}{-}\rangle_2 + \sin\theta\ |S_z{=}{-}\rangle_1\otimes|S_z{=}{+}\rangle_2$$
$$|A_2\rangle = \sin\theta\ |S_z{=}{+}\rangle_1\otimes|S_z{=}{-}\rangle_2 + \cos\theta\ |S_z{=}{+}\rangle_1\otimes|S_z{=}{-}\rangle_2$$

for some angle θ. Mathematically speaking, the two-particle spin states picked out by the biorthogonal decomposition lie in the subspace of Hilbert space spanned by $|S_z{=}{+}\rangle_1\otimes|S_z{=}{-}\rangle_2$ and $|S_z{=}{-}\rangle_1\otimes|S_z{=}{+}\rangle_2$. Physically speaking, we know this has to be the case, or else the particles would have nonzero probability of yielding the same outcomes (i.e., two ups or two downs) upon measurement of their z-spins.

If $|d_1| \neq |d_2|$, the biorthogonal decomposition of |intermediate> is unique. Where does this leave us? Formally, there exists a Hermitian operator A of which $|A_1\rangle$ and $|A_2\rangle$ are eigenstates. According to the modal interpretation, the observable A

---

[36]This does not mean that $|A_1\rangle = |S_n{=}{+}\rangle_1\otimes|S_n{=}{-}\rangle_2$ for some direction **n** that's very close to **z**. In general, $|A_1\rangle$ and $|A_2\rangle$ will be entangled states.

corresponding to A has a definite value. As just noted, the property of A-definiteness encodes, among other things, a perfect anticorrelation between the particles' z-components of spin. And A-definiteness is a holistic property, in the sense discussed above. The particles "have" a z-spin anticorrelation even though neither individual particle has a definite z-component of spin.

If $|d_1| = |d_2|$, then |intermediate> is a *non-unique* biorthogonal decomposition with respect to the subsystem under consideration, the spin characteristics of the two-particle system. Healey's modal interpretation would pick out as definite a degenerate observable A, where A is a projector onto the subspace spanned by $|S_z{=}{+}{>}_1 \otimes |S_z{=}{-}{>}_2$ and $|S_z{=}{-}{>}_1 \otimes |S_z{=}{+}{>}_2$. So, that degenerate A would encode the same holistic property of z-spin-anticorrelation just discussed.

Let me summarize the results so far. Initially, and also in the intermediate stage, neither particle has a definite z-component of spin. Before particle 1 passes through the magnet, the two-particle system has a definite value for $P_{|singlet>}$, which corresponds to perfect spin anticorrelations in all directions. By contrast, after particle 1 traverses the magnet, the two-particle system no longer has a definite value of $P_{|singlet>}$. Instead, it has a definite value of A, which corresponds to a perfect spin anticorrelation *in the z-direction only*. This modal value assignment reflects the fact that, if we measure the n-spins of both particles after particle 1 passes the z-aligned magnet, we won't always get opposite outcomes, unless **n=z**.

§4.5.2. Singlet-state: After measurement

Now I'll explore what properties become definite after particle 1 strikes one of the photographic plates. To do so, I must first write the updated quantum state. The particle's spin gets "disturbed" when it interacts with the electrons and other particles

comprising the plate. In quantum mechanical terms, this means that the spin of particle 1 becomes entangled with the spin (and orbital) angular momentum of the particles in the plate. To capture this fact, I'll write the final state of the photographic plates-*and their environment* as |R=up,ij>, where "i" denotes the z-spin of particle 1 before striking the plate, and "j" denotes the z-spin of particle 1 *after* interacting with the plate. So for instance, |R=up,+-> denotes the final state of the plates (and their environment) when particle 1, initially in state $|S_z=+>$, leaves a dot on the "up" plate and has its spin "flipped" to $|S_z=->$. Similarly, |R=down,++> denotes the case where particle 1, despite having initial z-spin up, nonetheless strikes the down plate (due to the spatial wavefunction tail of $|\phi_{up}>$) and does not have its spin flipped. Notice that I'm "building the environment" into these states, instead of writing the environmental states separately.

Let's think about the difference between |R=up,++> and |R=up,+->. In both cases, particle 1 leaves a dot on the "correct" plate, by which I mean the plate corresponding to the pre-measurement spin of particle 1. The only difference between these two states is whether the plate "flips" the spin of particle 1. This depends on microscopic interactions between particle 1 and the plate particles. So, we expect that |R=up,++> and |R=up,+-> differ microscopically *but not macroscopically*. Nonetheless, |R=up,++> and |R=up,+-> are orthogonal, because they correspond to states of different angular momentum. To see why, suppose a z-spin-up particle hits the upper plate. (In this mini-experiment, there's no particle 2.) Then the final state of the universe is a superposition of $|R=up,++>\otimes|S_z=+>$ and $|R=up,+->\otimes|S_z=->$. Because angular momentum is conserved during the interaction, these two branches of the superposition must have the same angular momentum. Since $|S_z=+>$ and $|S_z=->$ differ in angular momentum by $\hbar$, so must |R=up,++> and |R=up,+->. For instance, if the plate before the interaction had total

angular momentum 0, then |R=up,++> must have angular momentum 0, while |R=up,+-> must have angular momentum $+\hbar$.

By contrast, |R=up,++> and |R=up,--> are not necessarily orthogonal. Remember, |R=up,--> is the plates' state when an initially down particle "mistakenly" hits the up plate, and doesn't have its spin flipped. And |R=up,++> is the plates' state when an initially up particle "correctly" hits the up plate, and doesn't have its spin flipped. So, for both |R=up,++> and |R=up,-->, the plate's total angular momentum is unchanged during the interaction.

I'll now return to the two-particle EPR-type experiment discussed above, and write the state of the universe after particle 1 hits one of the plates, but before particle 2 interacts with anything. I'll use boldface coefficients to indicate "big" terms. The other terms are tiny, because they stem from wavefunction tails (e.g., an initially up particle hitting the down plate):

$$\begin{aligned} |\text{final}\rangle = \quad & |S_z{=}{-}\rangle_2\{\mathbf{c_{11}}|\mathbf{R{=}up,{+}{+}}\rangle\otimes|\mathbf{S_z{=}{+}}\rangle_1 + \mathbf{c_{12}}|\mathbf{R{=}up,{+}{-}}\rangle\otimes|\mathbf{S_z{=}{-}}\rangle_1 \\ & + d_{11}|R{=}down,{+}{+}\rangle\otimes|S_z{=}{+}\rangle_1 + d_{12}|R{=}down,{+}{-}\rangle\otimes|S_z{=}{-}\rangle_1\} \\ + \ & |S_z{=}{+}\rangle_2\{c_{21}|R{=}up,{-}{+}\rangle\otimes|S_z{=}{+}\rangle_1 + c_{22}|R{=}up,{-}{-}\rangle\otimes|S_z{=}{-}\rangle_1 \\ & + \mathbf{d_{21}}|\mathbf{R{=}down,{-}{+}}\rangle\otimes|\mathbf{S_z{=}{+}}\rangle_1 + \mathbf{d_{22}}|\mathbf{R{=}down,{-}{-}}\rangle\otimes|\mathbf{S_z{=}{-}}\rangle_1\}. \end{aligned}$$

Before looking at the possessed properties of the particles, let's first confirm that the pointer-reading is definite, i.e., that a dot *really is* on the upper plate or on the lower plate. We can't answer this question simply by looking at |final>, because I've built the environment into the pointer-reading states, instead of teasing them apart. But we know from earlier considerations that the |R=up,ij> states decohere with the |R=down,ij> states. Therefore, when we biorthogonally decompose |final> with respect to the pointer

reading states (i.e., the state of the photographic plates), a degenerate observable gets picked out that's very close to the observable corresponding to R=up and R=down states. In other words, the plates *have* a dot either on the up or on the down plate. This conclusion fails only if the relevant biorthogonal decomposition is degenerate, which will be the case if and only if R=up and R=down are *exactly* equally likely. If the measurement were ideal, this degeneracy would kick in, since particle 1 is "up" half the time and "down" half the time. But due to the wavefunction tails, obtaining a dot on the up vs. the down plate is *exactly* equally likely only if, for instance, the plates are exactly the same size and are placed in a precisely symmetric configuration with respect to the initial state of the particles. If the up plate is (say) an angstrom closer to the magnets than the down plate is, then it has a slightly higher chance of getting hit. The probability is actually 0 that the two plates have *exactly* the same probability of getting a dot. Ironically, the impossibility of performing a perfectly ideal measurement saves the modal interpretation from a basis-degeneracy disaster.

But because the "up" and "down" probabilities are so close, the relevant biorthogonal decomposition will be nearly degenerate. Normally, this would imply that the biorthogonally-selected basis is not even close to the desired pointer-reading basis. But here's where decoherence saves the day. As time passes, the environmental states corresponding to R=up and R=down become closer and closer to orthogonal. This ensures that, no matter how close to a degeneracy we're "standing," decoherence will eventually ensure that the biorthogonal decomposition picks out states close to |R=up> and |R=down>. Roughly put, unless there's an exact degeneracy, decoherence eventually "knocks" the photographic plates into a state close to |R=up> or |R=down>.

But this isn't good enough, if the pointer reading stays indefinite for a noticeable length of time. Intuitively, the dot had better appear as soon as we develop the

photographic plates! Fortunately, decoherence acts sufficiently quickly. See Bacciagaluppi and Hemmo (1995) for lots of formal details. For instance, suppose the R=up and R=down outcomes have probabilities that differ by only 1 part in 10^{30}. Since the dots are macroscopic, decoherence ensures that states very close to |R=up> and |R=down> get picked out in under a thousandth of a second. Decoherence can "overcome" even the most severe near-degeneracy.

So, if you accept the "closeness" arguments from section 4.4, the modal interpretation ensures that a system in state |final> has a definite pointer-reading, by which I mean a definite dot on one of the plates.

Given all that, let's return to questions of nonlocality and holism. In the modal interpretation, does this definite measurement result on particle 1 cause particle 2 to (nonlocally) acquire a definite z-component of spin, as would be the case in "wavefunction collapse" models? No. To see why not, notice that the quantum mechanical density operator describing particle 2 does not change when particle 1 gets measured. In state |singlet>, |intermediate>, or |final>, the reduced density operator of particle 2 is

$$\rho_2 = 2^{-1}\{|S_z{=}+\rangle\langle S_z{=}+| + |S_z{=}-\rangle\langle S_z{=}-|\},$$

which can be rewritten as $\rho_2 = 2^{-1}\{|S_n{=}+\rangle\langle S_n{=}+| + |S_n{=}-\rangle\langle S_n{=}-|\}$ for any direction **n**. This is just to say that the biorthogonal decomposition of |final> with respect to particle 2 is non-unique.[37] So, according to the modal interpretation, particle 2 does not *have* a definite value of S_n for any **n**. Measuring particle 1 does not bring into existence a

[37]A reduced density operator is diagonalized in terms of a unique basis if and only if the biorthogonal decomposition of the total quantum state uniquely picks out that same basis.

physical property associated with particle 2. In this sense, the modal interpretation does not violate EPR's locality requirement.[38]

Nonetheless, if we get a dot on the upper plate, we *know* that subsequent z-spin measurement of particle 2 will almost certainly yield "down." The modal interpretation must encode this fact, or else it's "incomplete" in some sense.[39] To see how the modal interpretation "completes" itself, it helps to regroup the terms in |final>:

$$
\begin{aligned}
|\text{final}\rangle = \quad & |S_z{=}+\rangle_1\{\mathbf{c_{11}|R{=}up,{+}{+}\rangle\otimes|S_z{=}-\rangle_2} + d_{11}|R{=}down,{+}{+}\rangle\otimes|S_z{=}-\rangle_2 + \\
& + c_{21}|R{=}up,{-}{+}\rangle\otimes|S_z{=}+\rangle_2 + \mathbf{d_{21}|R{=}down,{-}{+}\rangle\otimes|S_z{=}+\rangle_2}\} \\
+ \quad & |S_z{=}-\rangle_1\{\mathbf{c_{12}|R{=}up,{+}{-}\rangle\otimes|S_z{=}-\rangle_2} + d_{12}|R{=}down,{+}{-}\rangle\otimes|S_z{=}-\rangle_2 + \\
& + c_{22}|R{=}up,{-}{-}\rangle\otimes|S_z{=}+\rangle_2 + \mathbf{d_{22}|R{=}down,{-}{-}\rangle\otimes|S_z{=}+\rangle_2}\}.
\end{aligned}
$$

Since the four boldfaced terms are either precisely or almost precisely mutually orthogonal,[40] and the other terms contribute negligibly, this expansion of |final> is almost biorthogonal with respect to particle 1. Unless the coefficients happen to add up in *exactly* the right way, the biorthogonal decomposition of |final> with respect to particle 1 will be unique. Here's why. The density operator describing particle 1 is non-uniquely

---

[38]EPR-locality demands, roughly speaking, that nothing we do to particle 1 can instantaneously bring into existence an "element of reality" associated with particle 2.

[39]See Elby, Brown, and Foster (1993) for a detailed discussion of what "incomplete" means. We contrast "EPR-completeness" with "statistical completeness." In the present context, these fine distinctions aren't important.

Digression for EPR fans: Even when particle 1 yields "up," particle 2 does not have probability 1 of yielding "down," due to wavefunction tails. So, the physical property that encodes particle 2's near-certainty of yielding z-spin down does not meet EPR's sufficient condition for being an "element of reality." As just noted, according to the modal interpretation, this physical property--whatever it turns out to be--is not a physical property *of particle 2 per se.*

[40]Recall from above that |R=up,++> and |R=up,+-> are strictly orthogonal, due to angular momentum conservation. And |R=up,ij> is almost orthogonal to |R=down,ij>, due to decoherence.

diagonalizable only if $\rho_1 = 2^{-1}\{|S_z=+><S_z=+| + |S_z=-><S_z=-|\}$, which is the case (roughly speaking) only if particle 1 has an *exactly* 50% chance of finishing the measurement interaction with spin "up." This will be the case if, for instance, both plates have exactly the same probability of flipping the particle's spin during the measurement interaction. Epistemically, of course, these "spin flip" probabilities for the upper and lower plate are equal. But the objective quantum probabilities depend on the microstates of the photographic plates. For instance, suppose the upper plate is exactly the same as the lower plate, except that the upper plate contains *one* extra atom of impurity. That one-atom difference changes the interaction Hamiltonian between the plate and particle 1, and thereby changes the odds that particle 1's spin gets flipped. Of course, this difference in odds is unbelievably small. But as long as it's nonzero, the biorthogonal decomposition (and equivalently, the density matrix written as a "mixture") avoids degeneracy.

Because we're so close to a degeneracy, the definite value associated with particle 1 corresponds to an operator that might not even be close to S_z. But that's o.k. Because the measurement interaction "disturbs" the particle, we don't physically expect particle 1 to end up with a definite value of S_z. The key fact is that there's a definite value associated with particle 1, because |final> can be *uniquely* biorthogonally expanded in the form $\sum_i c_i|A_i>\otimes|B_i>$, where $|A_i>$ are orthogonal states of particle 1 and $|B_i>$ are orthogonal states of plates/environment/particle 2. According to the modal interpretation, there exists an operator **A** associated with particle 1, of which $|A_i>$ are eigenstates. The corresponding observable A has a definite value, which is just to say that particle 1 *has* the definite property corresponding to $|A_i>$ for some i. (Here, i ranges from 1 to 2.) Similarly, there exists an operator **B** of which $|B_i>$ are eigenstates; and the corresponding observable B has a definite value. As just hinted, this B-definiteness is a

holistic property of the plates, particle 2, and the environment. Let's explore the nature of this property.

Crucially, the $|B_i\rangle$ states will be superpositions of the four boldfaced terms, with only small contributions from the other terms. And all four of those boldfaced terms correspond to a perfect anticorrelation between the measurement outcome on particle 1 and the z-spin of particle 2. So, the $|B_i\rangle$ states are extremely "close" to states that encode this same anticorrelation. In other words, plates/environment/particle 2 possesses a definite holistic property corresponding to an almost perfect anticorrelation between R and S_z (for particle 2).

Let me summarize and clarify these results. Due to decoherence, an observable very close to the "pointer-reading" R has a definite value; there *really is* a dot on one plate or on the other plate. Furthermore, the plates/environment/particle 2 possesses a property corresponding to a nearly perfect anticorrelation between R and the z-spin of particle 2. Nonetheless, particle 2 considered as an individual system does *not* have a definite z-spin. However, since something close to R is definite, and since R is anticorrelated with the z-spin of particle 2, the equation of motion governing these modally-possessed values ensures that, *if* the z-spin of particle 2 (or something correlated with that spin) ever acquires a definite value, that value will with high probability be anticorrelated with R. In other words, if we ever measure particle 2, it will almost certainly produce a dot on the "correct" plate.

In brief: After particle 1 gets measured but before particle 2 gets measured, particle 2 doesn't have a definite z-component of spin. But the modal interpretation still encodes the fact that subsequent z-spin measurement of particle 2 will almost certainly yield the opposite result to that obtained on particle 1. The modal interpretation does so by (i) assigning a definite value to the particle 1 measurement outcome, (ii) assigning a definite

holistic property corresponding to n anticorrelation between the particle 1 measurement outcome and the z-spin of particle 2, and (iii) having an equation of motion ensure that this anticorrelation becomes actualized when particle 2's z-spin (or some observable correlated with it) becomes definite.

In this scheme, the Bell-nonlocal connection between the two wings of the experiment takes the form of a holistic property, not a classically "causal" connection. It's not that particle 1 or its measuring apparatus sends a "signal" to particle 2 or otherwise affects a property of particle 2. Rather, measuring particle 1 causes a holistic property of apparatus/particle 2 to evolve in such a way that the spin-anticorrelation will almost certainly be manifested if particle 2 undergoes measurement.

The spirit of this discussion agrees with Healey (1994). But Healey's discussion doesn't treat the near-degeneracy problem, nor does it allow the spin of particle 1 to be disturbed during measurement. As a result, he and I disagree about what observables have definite values at what times. Nonetheless, we agree about the central role played by holistic properties in mediating--indeed, *constituting*--the nonlocal connection. Precisely because the modal interpretation is one of the few viable interpretations of nonrelativistic QM, these insights about holism might apply in some veiled form to nature itself, even if QM itself or the modal interpretation turns out to be wrong.

CHAPTER 5: CAUSATION VS. HOLISM

5.1. INTRODUCTION

This chapter explores whether we can causally explain the quantum mechanical EPR correlations, under the assumption that relativistic quantum theory is fundamental.

Other papers addressing this topic, including Redhead (1992) and Elby (1992), typically set necessary conditions on causation, and then show these conditions to rule out a "causal" explanation of EPR. As Healey (1992) emphasizes, however, different philosophers deploy different conceptions of causation in different contexts. Even the most popular conditions on causation, such as Reichenbach's principle of the common cause, may fail for certain notions of causation. Therefore, causation no-go theorems in the style of Redhead or Elby fail to establish that *no* variety of causal explanation can account for the EPR correlations within a quantum framework.

In this paper, I tease apart three often-combined yet distinct notions of causation. According to "minimalist" causation, a causal relation is nothing more than a suitably-formulated lawlike dependence between events. "Generative" causation demands that causes generate (bring about) their effects. And "continuity" causation requires that causal connections be mediated by continuous processes.

To disentangle these notions of causation, I evaluate their commitment to two necessary conditions: (1) A Reichenbachian screening off requirement called Reich; and (2) "Causal Unidirectionality," which (roughly) requires effects not to cause their causes. I'll argue that only generative causation is committed to both Reich and Causal Unidirectionality. By contrast, a causal minimalist can sensibly renounce Causal Unidirectionality in the case of spacelike causation. Minimalist causation can also

abandon Reich, but only by renouncing a compelling intuition I'll discuss below. Continuity causation can renounce Causal Unidirectionality, by dropping the "standard" though often tacit assumption that causal processes correspond to physical processes involving transport of energy density, current density, or some other conserved quantity. (Continuity causation can also renounce Reich, but only by allowing non-Markovian processes.) The bulk of this paper explores the philosophical ramifications of renouncing Reich or Causal Unidirectionality, for generative, minimalist, and continuity causation.

Then, I prove that within the framework of relativistic quantum theory, we cannot causally explain the EPR correlations consistent with Reich, Causal Unidirectionality, and a symmetry requirement that applies to all explanations, both causal and noncausal. Therefore, generative causation cannot account for EPR. But certain flavors of minimalist or continuity causation *can* account for these nonlocal quantum correlations. I'll explore what these "causal" explanations of EPR could look like.

In summary, this paper attempts to pinpoint which notions of causality provide a framework in which we can explain the EPR correlations within the context of relativistic quantum theory. Armed with these results, philosophers can debate whether we should explain EPR causally, or abandon causality in favor of a new, perhaps holistic explanatory framework.

2. THE EPR CORRELATIONS

Before discussing different notions of causation, I briefly review Bohm's version of the Einstein-Podolsky-Rosen (EPR) correlations.

In a typical EPR thought-experiment, two electrons, prepared in the spin singlet state, leave their source and travel in opposite directions towards measuring apparatuses. Both apparatuses measure the same component of "spin," which is a particle's intrinsic angular momentum. The "A-wing" of the experiment refers to one of the apparatuses along with the electron it measures, while the "B-wing" refers to the other apparatus and the electron it measures.[41]

According to relativistic quantum theory, when the two electrons occupy a spatially symmetric spin singlet state, they display the following characteristics:

- In the rest frame of the source, the probability density of finding an electron at spacetime point (x,t) equals the probability density of finding an electron at spacetime point (-x,t). The probability distributions for velocity are also symmetric.
- The spin properties of the two electrons are equivalent. Formally, the same spin density operator describes both electrons.

Furthermore, if the two measurements occur at spacelike separation, they cannot be objectively time ordered; neither measurement happens "before" the other. So, in relativistic quantum theory, the spatially symmetric spin singlet state Ψ corresponds to a physically symmetric state of affairs. Therefore, if the measuring apparatuses occupy the same quantum state, then relativistic quantum theory describes the A-wing and B-wing of the experiment equivalently (before either measurement occurs). According to relativistic quantum theory, interchanging the A-wing with the B-wing would result in the *same* physical state of affairs. Succinctly, *the pre-measurement physical state of affairs is symmetric* with respect to A-wing $\leftrightarrow$ B-wing exchange.

---

[41]Don't take this wording to suggest that we can talk sensibly about "this electron" versus "that electron." For now, I'm just describing the experiment in rough terms.

Of course, other theories describe this experiment asymmetrically. For instance, in nonrelativistic quantum mechanics and in Bohm's hidden-variable theory (see Bohm and Hiley 1987), the existence of "absolute time" allows a time-ordering of the two-measurements. This time ordering breaks the symmetry between the two wings.[42] Here, however, I will focus on causal explanations within the framework of relativistic quantum theory. That is, I'll assume relativistic quantum theory gives a true account. This is, of course, highly unlikely; but the "true" theory might resemble relativistic quantum theory in the relevant ways, so that my philosophical analysis still applies.

Let $\underline{\varepsilon}_a$ ($\underline{\varepsilon}_b$) denote the event of the A-wing (B-wing) apparatus measuring an electron and yielding result α (β). For electrons in state Ψ, whenever $\alpha=\hbar/2$, $\beta=-\hbar/2$; and whenever $\alpha=-\hbar/2$, $\beta=+\hbar/2$. Importantly, Ψ does not screen off these experimentally-confirmed correlations, as we'll see below.

Later on, I'll discuss whether, within the context of relativistic quantum theory, we can "causally" explain the EPR correlations. But first, I must outline some different notions of causation.

3. THREE NOTIONS OF CAUSATION

In this section, I review three intuitive, widely-held conceptions of causation. "Minimalist causation" asserts that a causal relation is nothing more than a lawlike dependence between cause and effect. "Generative causation" insists that causes *bring about* their effects. And "continuity causation" asserts that continuous processes mediate causal connections.

---

[42]Except when the two measurements occur simultaneously, which happens in a zero-measure subset of cases.

Of course, most theories of causation lean on intuitions drawn from more than one of the above. But to explore which notions of causality provide a framework in which we can explain EPR, I must *disentangle* these three notions. Given this goal, I will *not* discuss all of the objections raised against these three notions. For my arguments to get off the ground, I need to assume only that the three conceptions of causation, when properly fleshed out, could be rendered reasonably coherent.

These notions of causation agree that causal relations are relations between events. I will not address alternatives.

3.1. Minimalist causation. This viewpoint carries less ontological baggage than other varieties of causation do. According to causal minimalists, if a suitably formalized lawlike dependence holds between two events, then the events are causally related (or else jointly caused by a common cause), *simply by virtue of the lawlike dependence.* If *a* causes *b*, it's not because a physical process connects *a* to *b*, and it's not because *a* "brings about" *b* in some independent sense. Rather, it's because *a* and *b* satisfy formal relations encoding their lawlike dependence. Proponents of causal minimalism need not rely on potentially murky metaphysical constructions such as "continuous processes" or other causal mechanisms.

As an example, consider a standard "regularity" view of causation, according to which if $p(b|a)>p(b)$ is lawlike[43] then either *a* and *b* are directly causally connected, or else a common cause *c* is connected to both *a* and *b*. (Here, $p(x)$ denotes the objective probability that event *x* occurs, while $p(y|x)$ is the probability of *y* given *x*.) In this

[43]Delineating which correlations are lawlike and which are coincidental turns out to be notoriously difficult. In this paper, however, I can sidestep the issue, because I'm exploring the possibility of causal explanation *within the framework of relativistic quantum theory*. Therefore, for my purposes, a correlation is lawlike if predicted by relativistic quantum theory. These are the only correlations I address.

framework, the causal connections between events exist *by virtue of* these lawlike correlations, not by virtue of some independent metaphysical connection between the events.

Although Lewis's (1986) theory of causation differs markedly from regularity causation, Lewis-causation is also a variety of causal minimalism. According to Lewis, two events are causally connected if the propositions corresponding to the occurrence of those events, and the propositions corresponding to the non-occurrence of those events, satisfy a certain set of counterfactual relations. For my purposes, the details of Lewis's program aren't important. What's important is that, in Lewis's scheme, events are causally related *because* certain counterfactual statements are true, not because an independent physical or metaphysical connection links the events. Indeed, Lewis rejects talk of causal mechanisms as superfluous.

Despite the differences between competing versions of minimalist causation, all of them flesh out the same intuition: A causal relation is nothing more than a properly-formalized assertion that the events non-coincidentally, and perhaps even necessarily, tend to occur "together."

3.2. Generative causation. According to generative causation, *a* is a partial cause of *b* just in case *a* helps to *bring about b*..

The "generation" relation is stronger than a mere affirmation that *a* and *b* are correlated, even if the correlation supports counterfactuals. I can't (and therefore won't) formalize or explicate "generation." Rather, I'll treat it as a pre-systematic relation rooted in our causal intuitions.

Generative causation does not rule out superluminally-mediated causal connections, action at a distance, or even backwards-in-time causation. Generative causation

demands only that causes *bring about* their effects in some strong sense. Indeed, the central intuition underlying generative causation demands that all events, except the Big Bang, be brought about by other events. (I mean "event" in its broadest sense, as including the physical state of affairs in a spacetime region.) Presumably, several different formulations of causation fit wholly or partly into a generative-causal framework.

Before continuing, I must acknowledge a criticism that threatens the very notion of generative causation. One could claim that, upon closer examination, the distinction between generation and lawlike dependence breaks down, especially when the dependence supports counterfactuals. Here's the argument: Suppose that b necessarily occurs after a, and necessarily does not happen at any other time. Suppose also that no common cause of a and b exists. Then a "brings about" b in the sense that, when a happens, b must follow. What else could we possibly mean when we say a brings about b? Put another way, how can we coherently claim that a does not bring about b, given that b necessarily follows a and never happens at any other time (and given that no common cause exists)? According to this argument, we can ascribe no meaning to the "generation" relation that goes beyond the lawlike dependence between the events. Therefore, generative causation reduces to a version of minimalist causation.

In section 4.3, I'll show how we can tease apart generative causation from minimalist causation. Before tackling this issue, however, I'll introduce the third notion of causation addressed in this paper.

3.3. Continuity causation. Some philosophers, including Salmon (1984), argue that events are causally connected just in case they are connected by the right kind of continuous process. What counts as a "continuous process" varies by philosopher.

Indeed, the fuzzy concept of "process," notoriously difficult to define, threatens the coherence of continuity causation. But let's assume continuity causation can be rendered coherent.

As an example of continuity causation, consider a television set. Pressing the remote control button *causes* my TV to switch on, because continuous processes corresponding to propagation of electromagnetic radiation, flow of electrons, and so on, connect the button-pressing to the switching-on. Usually in this framework, the continuous causal process corresponds to a *physical* process involving energy-momentum transfer. This is true even in quantum mechanics, because a propagating wavefunction usually "carries" energy-momentum density. As Healey (personal communication) points out, calling wavefunction propagation a "process" makes sense only if we interpret the wavefunction non-instrumentally, as somehow coding real physical properties of the system.

According to continuity causation, an unmediated nonlocal correlation either doesn't exist or doesn't correspond to a causal connection. For instance, if *a* and *b* are correlated even though no continuous processes connect the events to each other or to common causes, then the correlation is noncausal.

In section 4.3, looking at Healey's explanation of EPR, we'll examine a continuous process that some philosophers might hesitate to call causal, and we'll explore the intuitions underlying this hesitation.

3.4. Wrap-up. A typical classical causal explanation invokes intuitions drawn from both generative causation and continuity causation. We often say that a cause brings about its effect *via* a continuous causal process. Indeed, causal intuitions might tempt us to assert that *a* generates *b* just in case the right kind of continuous process connects *a* to *b*. The

resulting hybrid notion of causation functions well in a "classical" nonrelativistic or relativistic universe. That's because classical correlations almost always result from direct contact action or from propagation of energy-momentum, both of which seem intuitively "generative." By contrast, nonlocal quantum correlations force us to choose which of our causal intuitions to retain and which to abandon. For this reason, we must further disentangle minimalist causation from generative causation from continuity causation. I'll now try to accomplish precisely that.

4. NECESSARY CONDITIONS ON CAUSATION

This section discusses two popular necessary conditions on causation: a Reichenbach-inspired screening-off requirement called Reich; and Causal Unidirectionality, the requirement that effects not cause their causes. Specifically, we'll explore which of the three notions of causation introduced above must obey these necessary conditions. Doing so will drive wedges between the different conceptions of causation. I'll also introduce Explanatory Symmetry, which requires a causal explanation to mirror any symmetries inherent in the underlying physical description of the phenomena.

As proven in section 5, any formulation of causation obeying Reich, Causal Unidirectionality, and Explanatory Symmetry *cannot* provide a causal explanation of the EPR experiment, within the framework of relativistic quantum theory. This result increases the importance of deciding which notions of causation must obey those conditions.

4.1. Reich. Reich demands that causes, when taken together, probabilistically *screen off* their effects from each other and from other events. If C is the set of *all* partial causes of some event e, then Reich insists that given C, no other event is probabilistically relevant to e, except effects of e:

Reich: C is *all* the partial causes of e only if $p(e|\, C,w) = p(e|\, C)$ for all w (besides e and its effects).

To explore whether minimalist causation must endorse Reich, let's jump right to the EPR correlations. Consider a standard EPR experiment in which the A-wing measurement occurs absolutely *before* the B-wing measurement. As mentioned above, $p(\underline{\varepsilon}_b \mid \underline{\varepsilon}_a, \Psi) \neq p(\underline{\varepsilon}_b \mid \Psi)$. Furthermore, since we're assuming that relativistic quantum theory holds, the correlation between $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$ is sufficiently lawlike and necessary to satisfy the formal "lawlike dependence" relations posed by any reasonable version of minimalist causation. Therefore, according to minimalist causation, *simply by virtue of that lawlike dependence,* either

(i) $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$ are directly causally connected; or else

(ii) $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$ are not causally connected, but there exists a common cause,

presumably Ψ (or Ψ in conjunction with other events).

Under choice (i), Reich holds. Under choice (ii), Reich fails, because the common cause doesn't screen off its effects from each other. This is true even if we build into Ψ the pre-measurement state of the entire universe.

Strictly speaking, minimalist causation is compatible with choice (ii), failure of Reich. But the spirit behind minimalist causation strongly motivates us to choose (i). Before discussing this motivation in general, let me illustrate it with a specific case, namely Lewis's minimalist-causal framework. According to Lewis's rules, Ψ is not the cause of $\underline{\varepsilon}_b$, because $\underline{\varepsilon}_b$ isn't necessitated by Ψ. But given Ψ, $\underline{\varepsilon}_a$ *is* the cause of $\underline{\varepsilon}_b$; given the state preparation Ψ, a down outcome occurs on the B-wing just in case an up outcome happens on the A-wing. See Butterfield's (1992) article, "David Lewis meets John Bell."

The following argument helps to motivate Lewis's (and other causal minimalists') choice of (i), and more generally, their endorsement of Reich. The state preparation alone does not determine which B-wing outcome will occur, or even which B-wing outcome is most probable. The B-wing result depends also on the A-wing outcome. Since $\underline{\varepsilon}_b$ irreducibly depends on $\underline{\varepsilon}_a$, even when all other factors are taken into account, and since causal relations are nothing more than appropriately-formulated lawlike dependencies, $\underline{\varepsilon}_a$ is a (partial) cause of $\underline{\varepsilon}_b$, exactly as Reich demands. This same intuitive argument, modified and augmented, applies to regularity causation. Put another way: Since causal connections *are* lawlike dependencies (or at least, correspond very closely to lawlike dependencies), each "independent" lawlike dependency should correspond to a separate causal connection. This is exactly what Reich demands.

For suppose Reich fails: C is the whole cause of e even though $p(e|C,w) \neq p(e|C)$. Then, even though w is lawlike correlated with e, and even though this correlation isn't screened off by other correlations, w is not causally connected to e. Nonetheless, C is causally connected to e. In other words, if we consider two correlations, both of which (I'll assume) satisfy the same formal rules of lawlike dependency and both of which are "independent" in the same sense, one of them might not correspond to a "causal"

connection even though the other one does. This seems unacceptable, given the causal minimalist's insistence that any lawlike dependency satisfying certain formal conditions *automatically* corresponds to causation (either a direct causal connection or else mutual dependence on a common cause).

In brief, a compelling minimalist-causal intuition motivates Reich: Each *independent* lawlike dependence in nature corresponds to a separate causal relation, so that no lawlike dependence is left "unexplained." A causal minimalist can reject Reich only at the expense of renouncing this intuition.

Crucially, the above arguments, and the conclusion just stated, apply equally well when $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$ are spacelike separated. In that case, *other* considerations may push us to deny a causal connection between $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$, and hence, to renounce Reich. But as we'll see below, those other considerations stem from continuity-causal intuitions, *not* from the minimalist-causal intuitions. For this reason, abandoning Reich in a purely minimalist-causal framework weakens the appeal of the resulting causal explanations.

4.2. Reich, generative causation, and continuity causation. So far, I've examined to what extent minimalist causation is committed to Reich. Now I'll discuss whether generative causation must endorse Reich.

Since generation is a stronger relation than lawlike correlation, it is *logically* possible for a lawlike correlation to result neither from direct causation nor from common causes. Such a correlation would violate Reich. But generative causation rules out Reich violation as *physically* (though not logically) impossible. Here's why:

According to the central generative-causal intuition, all lawlike correlations in the world stem from an intricate web of generative-causal connections. This intuition implies Reich. To see why, visualize each generative-causal connection (i.e., each

"bringing about") as a strand connecting two events. Since in generative causation all lawlike dependencies must be "explained" by this web, it follows that if *w* and *e* are lawlike correlated, then either a strand runs directly from *w* to *e*, or else there exists some common cause that sends strands to both w and *e* . But suppose Reich fails: p(*e*| *C*,*w*)>p(*e*| *C*), where *w* is not an effect of *e*, and where *C* is supposedly the whole cause of *e*. Since *C* is the whole cause of *e*, no strand runs from *w* to *e*, either directly or via a common cause. In other words, the lawlike dependence of *e* on *w* isn't explained by the generative web. As just noted, this contradicts the central intuition behind generative causation. Rejecting this intuition is tantamount to rejecting generative causation. For this reason, generative causation is committed to Reich as a necessary condition.

Notice how this argument resembles the minimalist-causal justification of Reich discussed above. A causal minimalist can (at great cost) reject the intuition that independent correlations warrant separate causal relations. But a generative causation advocate *cannot* reject this intuition, because each independent correlation must be "brought about," i.e., must correspond to a separate strand (or set of strands) in the generative web.

How does continuity causation fare with respect to Reich? Many advocates of continuity causation, including Salmon (1984), endorse a screening-off requirement. For concreteness, suppose a continuous process connects *c* (in spacetime region R_c) to *e* (in spacetime region R_e); and suppose no other events or processes are causally relevant to *e*. Under most formulations of continuity causation, the "state" *d* of the causal process in spacetime region R_d, where R_d is spatiotemporally between R_c and R_e, screens off *e* from *c*. In other words, given *d*, the probability of *e* does not depend on whether *c* occurred:

$$p(e|d,c) = p(e|d,\sim c),$$

where ~c denotes the non-occurrence of *c*. More generally, each stage of the continuous causal process screens off the effect from preceding stages.

Despite the traditional incorporation of screening off into continuity causation, I now argue (following Cartwright and Jones 1992) that continuity causation need not obey Reich.

Suppose the $c \rightarrow d \rightarrow e$ process described above is non-Markovian, so that $p(e|d) \neq p(e|d,c)$. This inequality holds *not* because an independent continuous process links *c* to *e* (without passing through *d*), and not because the "process" under consideration is really the result of multiple intertwining processes, but simply because later stages in the continuous causal process do not screen off earlier stages. In this non-Markovian case, Reich suggests that we call *c* and *d separate* causes of *e*. But continuity-causal intuitions lead us to assert that *c* and *d* are *not* separate causes of *e*; instead, *c* and *d* are merely different stages of the *same* continuous causal process leading to *e*. Here's my point: Given non-Markovian processes, a continuity causation advocate can sensibly renounce Reich as a necessary condition.

Continuity causation advocates such as Salmon could defend Reich by denying the possibility of non-Markovian processes. To do so, they could emphasize that modern physical theories, including relativistic quantum theory, rule out non-Markovian fundamental processes. More precisely, according to *all* physical theories (that I know of) formulated in the past 300 years, the evolution of a system's fully-specified physical state S is strictly Markovian: S(t) screens off S(t+dt) from all previous states of the system. In words, a system does not "remember" its past state, *except* insofar as those memories are stored in the current state. Physical intuitions suggest that future physical

theories will also be Markovian; see Elby and Foster (1992). For this reason, a continuity causation advocate can deny the existence of non-Markovian "processes," and thereby defend Reich, by claiming that a "continuous causal process" corresponds to a physical processes, which by definition supervenes on the evolution of the fully-specified physical state. It's unclear, however, whether *a priori* philosophical considerations, other than Markovian physical intuitions, give us reason to rule out non-Markovian processes.

In summary: Minimalist causation can give up Reich, but only by renouncing the intuition that each *independent* lawlike dependence between events correspond to a separate causal relation. Generative causation cannot renounce this intuition, because each independent lawlike correlation must be "brought about." In continuity causation, Reich pops out as a compelling necessary condition only if we insist that "continuous causal processes" correspond to physical processes. Within the framework of relativistic quantum theory, this correspondence ensures that causal processes will be Markovian, because physical processes are Markovian.

4.3. Causal Unidirectionality. My next allegedly necessary condition on causation is

Causal Unidirectionality: If two events are causally connected, then one "causes" the other, but not vice versa. Therefore, if a is a cause of b, then b cannot be a cause of a (unless a and b are part of a closed timelike loop).

This condition asserts that causal connections consist of "cause" and "effect"; and that effects cannot cause their causes. Within a causal explanation of events, if a somewhere

functions as a cause of *b*, then nowhere in the explanation may *b* function as a cause of *a*, unless *a* and *b* are part of a closed timelike loop.

Causal Unidirectionality as formulated here is absolute, not observer-relative. To see what this means, suppose *a* and *b* are spacelike separated. Observer A (B) inhabits a reference frame in which *a* (*b*) occurs first. In some cases, we might be tempted to claim that *a* causes *b* for observer A, while *b* causes *a* for observer B. But Causal Unidirectionality rules out such an explanation. Roughly speaking, Causal Unidirectionality requires that the direction of causation be an unambiguous fact about the world, not an observer-relative vestige of our causal explanations.

Generative causation must endorse Causal Unidirectionality. Intuitively, the generation relation is intrinsically asymmetric, even when the events can't be time ordered: If *a* generates *b*, then it makes no sense to say that *b* generates *a* .[44] Because a symmetric "bringing about" relation would violate our deepest intuitions about generation, any resulting "explanation" would fall outside the framework of generative causation. For this reason, Causal Unidirectionality is a reasonable necessary condition to place on generative causation.

I'll now show that minimalist causation need not obey Causal Unidirectionality. To do so, I'll first argue that minimalist causation need not rule out spacelike causation. For spacelike causation, I'll then argue, minimalist-causal intuitions do not promote Causal Unidirectionality.

---

[44]At first glance, this conclusion seems fishy in the context of a closed timelike loop. You might want to say that "*a* causes *b* causes *a* causes *b*..." But how did the loop itself come to be? This question illustrates the difficulty of retaining a generative causal framework when there's closed timelike loops. Fortunately, I can sidestep this tricky issue by confining my attention to relativistic quantum theory in *our* universe, where closed timelike loops are impossible. In the absence of such loops, the "generation" relation is intrinsically asymmetric.

Almost all versions of minimalist causation either postulate or imply that, if two causally connected events are timelike separated, then the earlier event is the cause, while the later event is the effect. But suppose *a* and *b* are spacelike separated. A causal minimalist could simply rule out nonlocal causation. But this move violates the spirit of minimalist causation. In a minimalist framework, a causal connection does not correspond to a fancy metaphysical construction. Rather, it expresses a properly-formulated lawlike dependence between two events. If *a* and *b* are lawlike correlated, and this correlation is "independent" from the correlations between those events and their other causes, then minimalist-causal intuition strongly suggests that a separate causal connection links *a* and *b*. In subsection 4.1, I argued this point more fully, to show that causal minimalists have good reason to endorse Reich. As we'll see below, our intuitions against spacelike causation come from continuity causation. So, "pure" minimalist causation has no reason to rule out nonlocal (spacelike) causal connections.

And given a causal connection between spacelike separated *a* and *b*, a "pure" causal minimalist has no incentive to deny that the events mutually cause each other. Or better yet, the causal minimalist can dispense with asymmetry-connoting talk of "cause and effect" in favor of a symmetry-connoting talk of a "directionless causal link." For if neither event precedes the other, then what motivation remains for calling one event the "cause" and the other event the "effect"? By accurately mirroring the directionless (symmetrical) lawlike dependence between the events, the directionless causal link avoids introducing excess ontological baggage. And this avoidance of superfluous metaphysics is a cornerstone of minimalist causation. So, minimalist causation is not committed to Causal Unidirectionality.

In response, a causal minimalist who likes Causal Unidirectionality could argue that it's nonsensical to say two events mutually cause each other in any sense. But as we saw

above, this argument rests on a *generative-causal* intuition about the asymmetry of the generation relation. Here's my point: If we really view causal relations as corresponding to nothing more than lawlike dependency, then we have no *a priori* incentive to deny a symmetric causal link between *a* and *b*. Nor do we have reason to deny that *a* causes *b* and *b* causes *a*.

Of course, a causal minimalist may adhere to Causal Unidirectionality, even for spacelike causation. This fact does not threaten my argument that a causal minimalist can *choose* whether to adopt Causal Unidirectionality as a necessary condition. Raw minimalist causation is not *committed* to Causal Unidirectionality.

For this reason, Causal Unidirectionality functions as a wedge we can drive between generative causation and minimalist causation. Recall from subsection 3.2 that according to some philosophers, the generation relation, when stripped of its rhetorical gloss, amounts to nothing more than a necessary lawlike dependence between two events; and hence generative causation reduces to minimalist causation. But generative causation, unlike minimalist causation, is committed to Causal Unidirectionality. For example, suppose that for spacelike separated *a* and *b*, the (non)occurrence of *a* necessitates the (non)occurrence of *b*, and vice versa. Here, the "necessitation" relations are symmetric. The generation relation, by contrast, is intrinsically asymmetric. For this reason, "generation" does not reduce to a lawlike dependence (such as necessitation). In brief, Causal Unidirectionality teases out a distinction between minimalist causation and generative causation, showing that distinction to be more than merely semantic.

In reply, as Paul Teller (personal communication) points out, a detractor of generative causation could argue as follows: Let m'-causation denote a variant of minimalist causation that adopts Causal Unidirectionality as a necessary condition. Then generative causation, when stripped of metaphysical fluff, reduces to m'-causation.

I have two responses to this. First, as argued above, Causal Unidirectionality is motivated by generative-causal intuitions, not by purely minimalist-causal intuitions. Therefore, m'-causation is two-faced: It relies on generative-causal intuitions to motivate Causal Unidirectionality, and then dismisses as nonsensical the metaphysics underlying those intuitions. Or, if the m'-causation advocate denies relying upon generative-causal intuitions, but fails to provide a purely minimalist-causal motivation for Causal Unidirectionality, then m'-causation is *ad hoc,* in that it contains an unmotivated rule. For these reasons, m'-causation isn't a palatable alternative to generative causation.

But even if generative causation ultimately reduces to a less metaphysically-loaded version of causation, this paper still makes a worthwhile argument. In section 5, I prove that *any* theory of causation committed to Causal Unidirectionality, Reich, and Explanatory Symmetry cannot account for the EPR correlations within the framework of relativistic quantum theory. In my view, a popular and intuitive conception called generative causation satisfies these conditions. If generative causation reduces to something else, then my proof applies to that "something else," whatever it is.

So much for generative causation vs. minimalist causation. Let's now explore whether continuity causation must obey Causal Unidirectionality.

In standard continuity-causal explanations, a continuous causal process corresponds closely to a *physical* process involving transfer of some conserved quantity, such as energy, angular momentum, and baryon number. This is true even in relativistic quantum mechanics, because a propagating wavefunction carries energy density, current density, etc. According to relativistic theories, both classical and quantum, transfer of a conserved quantity cannot exceed the speed of light.[45] Therefore, such processes

---

[45]Specifically, the "center" of a system's energy density, or angular-momentum density, or baryon-number density, etc., cannot exceed the speed of light, and hence cannot connect two spacelike separated events.

propagate forward in time. For this reason, if two events are causally connected by such a process, the earlier (later) one can unproblematically be called the cause (effect). Therefore, Causal Unidirectionality automatically holds in standard continuity-causal explanations.

But the nonseparability of entangled wavefunctions in quantum mechanics may open the door to "nonstandard" continuous processes, ones that don't correspond to transport of a conserved quantity. Some such processes may violate Causal Unidirectionality.

As a detailed example of a purported Unidirectionality-violating "continuous process," consider a crucial component of Richard Healey's (1992) explanation of the EPR experiment. His explanation is embedded within relativistic quantum theory *without wavefunction collapse*. Consider the case where $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$ occur at spacelike separation; neither measurement happens (absolutely) before the other. Suppose Ms. A observes the experiment from a reference frame in which $\underline{\varepsilon}_a$ happens before $\underline{\varepsilon}_b$. According to Ms. A, $\underline{\varepsilon}_a$ happens at time t_0 and $\underline{\varepsilon}_b$ happens at t_1, where $t_1 > t_0$. At t_0, the two-electron wavefunction becomes entangled with the A-wing apparatus. At t_1, this entangled wavefunction becomes further entangled with the B-wing apparatus. Relativistic quantum theory describes how, from Ms. A's perspective, the wavefunction evolves between t_0 and t_1. Crucially, this entangled wavefunction is nonseparable; the A-wing and B-wing of the experiment are holistically connected. The wavefunction evolution between t_0 and t_1 is *continuous*, in that the wavefunction at time t+dt differs only infinitesimally from the wavefunction at time t, for all t between t_0 and t_1. And this continuous wavefunction evolution connects $\underline{\varepsilon}_a$ to $\underline{\varepsilon}_b$ in the following sense: At t_0, "part" of the nonseparable wavefunction is localized at the A-wing measuring device. The wavefunction evolves between t_0 and t_1 such that at t_1, "part" of the wavefunction is

localized at the B-wing apparatus.[46] So, the wavefunction evolution connects $\underline{\varepsilon}_a$ (at t_0) to $\underline{\varepsilon}_b$ (at t_1).

Similarly, Mr. B, who observes the experiment from a reference frame in which $\underline{\varepsilon}_b$ precedes $\underline{\varepsilon}_a$, can give a description of how the nonseparable wavefunction continuously evolves so as to connect $\underline{\varepsilon}_b$ (at Mr. B's t_0) to $\underline{\varepsilon}_a$ (at Mr. B's t_1).

According to Healey, the continuous process linking $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$ corresponds to the *conjunction* of the wavefunction-evolution descriptions given by Ms. A and Mr. B. This process is "continuous" and "connecting" in that, according to any observer,[47] the nonseparable wavefunction evolves continuously between when the "first" and "second" measurements occur; and the wavefunction evolution connects those two measurements in the sense described above.

This continuity-causal partial explanation of EPR violates Causal Unidirectionality. Ms. A would say that $\underline{\varepsilon}_a$ is a partial cause of $\underline{\varepsilon}_b$, while Mr. B would claim $\underline{\varepsilon}_b$ is a partial cause of $\underline{\varepsilon}_a$. In Healey's explanatory framework, neither observer is "absolutely" right; the explanation deploys $\underline{\varepsilon}_a$ as both a cause of $\underline{\varepsilon}_b$ and an effect of $\underline{\varepsilon}_b$. (Alternatively, a continuity causation advocate could claim that talk of causes and effects makes no sense; we can speak only of the continuous process that mediates the causal connection between events.) Either way, Healey's explanation violates Causal Unidirectionality, which requires $\underline{\varepsilon}_a$ to function in the explanation as either the cause or the effect of $\underline{\varepsilon}_b$, but not both. Interestingly, however, Healey's explanation obeys an observer-relative version of Causal Unidirectionality:

---

[46]If you consider talk of "parts" to be out of place for entangled wavefunctions, then here's what I mean: At t_0, the electron probability density at the A-wing apparatus is significant; and at t_1, the electron probability density at the B-wing apparatus is significant.

[47]It's not clear how Healey's scheme treats an observer for whom the two measurements occur exactly simultaneously. But this might not be too important; see footnote 2 above.

Observer-relative Causal Unidirectionality: If two events are causally connected, then any particular observer will describe exactly one event as the "cause." Therefore, if a particular observer claims that *a* is a cause of *b*, then she cannot also claim that *b* is a cause of *a* (unless *a* and *b* are part of a closed timelike loop).

A "particular observer" differs from the "detached explainer," whose overall explanation of the events must subsume the descriptions given by different observers.

A continuity causation advocate who endorses Causal Unidirectionality will deny that evolution of the nonseparable wavefunction counts as a continuous causal process linking the two measurement outcomes. This advocate could argue that unless a single identifiable *part* of the wavefunction actually propagates from $\underline{\varepsilon}_a$ to $\underline{\varepsilon}_b$ (or vice versa), we cannot say the wavefunction evolution causally *connects* $\underline{\varepsilon}_a$ to $\underline{\varepsilon}_b$. Healey could reply that since the nonseparable wavefunction is holistic, we cannot sensibly talk about individuated parts of the wavefunction. The Causal Unidirectionality advocate's best response, I think, is to fall back on the "standard" continuity-causal requirement that a causal process correspond to a physical process involving transfer of some conserved quantity. By this argument, since no probability current density (and hence no energy density) flows from $\underline{\varepsilon}_a$ to $\underline{\varepsilon}_b$, the continuous process does not connect $\underline{\varepsilon}_a$ to $\underline{\varepsilon}_b$.

The above paragraph raises a crucial issue: To deny that unusual Unidirectionality-violating processes (such as Healey's) are "causal," a continuity causation advocate has little choice but to fall back on the requirement that a continuous causal process correspond to a physical process involving transfer of a conserved quantity from cause to effect. Therefore, we must exhume the intuitions underlying this standard requirement.

Intuitively, quantities such as energy and angular momentum are "active": When transferred to an object, they change the object's properties. In a causal process, if the relevant "effect" follows physically from these changed properties, then the transfer of energy or angular momentum "brings about" the effect, in some strong intuitive sense. By contrast, Healey's nonseparable wavefunction evolution does not seem to carry a generative agent from $\underline{\varepsilon}_a$ to $\underline{\varepsilon}_b$. Instead, the nonseparable wavefunction mediates a holistic, non-generative connection between $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$. To deny that seemingly "passive" connections between events (such as Healey's nonseparable wavefunction evolution) count as "causal," a continuity causation advocate must insist that a causal process carry something "active" (generative) from cause to effect. So, to escape Healey's conclusion that certain continuous processes violate Causal Unidirectionality, a continuity causation advocate must smuggle in generative-causal intuitions.

Healey's explanation of EPR is unusual in that it employs continuity causation *almost completely disentangled from the generative-causal intuitions usually present in continuity-causal explanations*. The resulting "99% pure"[48] continuity causation violates Causal Unidirectionality. Of course, some philosophers will argue that "pure" continuity causation fails to provide intuitively pleasing explanations. I'll address this point in section 6. For now, let me reiterate my goal of disentangling generative causation from continuity causation from minimalist causation, to see how each notion of causation fares with respect to EPR. Healey proves that we *can* causally explain EPR, if we employ continuity causation stripped of essentially all generative-causal intuition.

In summary, although most theories of causation assume or imply Causal Unidirectionality, only generative causation is committed to this necessary condition.

---

[48]Healey (personal communication) notes that his explanation may harbor the consequence of a residual g-causal intuition: Within a given reference frame, we can still talk of the "cause" and the "effect"; and cause precedes effect. That is, Observer-relative Causal Unidirectionality holds.

When we strip continuity causation of all generative-causal intuitions, thereby allowing Healey-style "passive" processes to count as causal, Causal Unidirectionality can fail. To defend Causal Unidirectionality, a continuity causation advocate can impose the standard requirement that causal processes correspond to physical processes involving transfer of conserved quantities. As just shown, however, generative-causal intuitions motivate this standard requirement. Furthermore, for spacelike separated events, minimalist-causal intuitions *alone* do not rule out a symmetric, directionless causal link. According to minimalist causation, a causal relation corresponds closely to a lawlike dependence between events, with no superfluous metaphysics. Therefore, a completely symmetric lawlike dependence could correspond to a symmetric causal connection.

4.3. Explanatory symmetry. My third necessary condition on causation requires causal explanations to mesh with physical descriptions:

Explanatory Symmetry: Suppose that a physical description **D** of some phenomena incorporates a fundamental symmetry. Let **E** be an explanation of the phenomena. If **E** takes **D** to be a "complete" physical description of the phenomena in question, then **E** must reflect the symmetry of **D**.

A physical description is "complete" only if the theory on which it is based is fundamental, and only if the description is as fine-grained as the theory allows.

Because Explanatory Symmetry is technically ambiguous (e.g., I don't supply a sufficient condition for the "completeness" of a physical description, or a definition of "fundamental" symmetry), it must serve more as a guiding principle than as a formal necessary condition. The "version" of this principle I'll invoke later applies to EPR:

EPR Explanatory Symmetry: If our physical description of the EPR experiment, assumed to be complete, is physically symmetric under A-wing ↔ B-wing exchange, then our corresponding explanation must not introduce an asymmetry between the two wings.

Relativistic quantum theory provides a description of the EPR experiment that is symmetric under A-wing ↔ B-wing exchange. Therefore, if we explain EPR within the context of relativistic quantum theory, EPR Explanatory Symmetry implies the following:

$$\text{If } \underline{\varepsilon}_a \text{ is a partial cause of } \underline{\varepsilon}_b, \text{ then } \underline{\varepsilon}_b \text{ is a partial cause of } \underline{\varepsilon}_a.$$

Explanatory Symmetry, I now argue, is a reasonable constraint on *all* varieties of causal and noncausal explanation. An explanation should do more than reiterate our bare-bones physical description of the events. The explanation should complement or flesh out those physical details, thereby helping us to "understand" the phenomena more deeply. Put another way, our metaphysical description of events, which may include causal connections, should combine with our physical description to provide a unified, coherent "picture" of the phenomena.

These considerations immediately motivate Explanatory Symmetry. When Explanatory Symmetry fails, our physical and metaphysical descriptions "disagree" about whether the phenomena are symmetric. Therefore, we cannot unify those physical and metaphysical components into a pleasing, coherent picture of what's going on.

Let's specialize these considerations to the EPR correlations. According to relativistic quantum theory, the A-wing and B-wing are physically equivalent; exchanging the two wings would make no physical difference.[49] This physical description strongly suggests that neither wing is "special." But if our explanation violates EPR Explanatory Symmetry, then one wing of the experiment gets singled out. For instance, if we claim that $\underline{\varepsilon}_a$ causes $\underline{\varepsilon}_b$, but not vice versa, then we've picked out the A-wing as causally "special." Such an explanation, by clashing with our symmetric physical description, fails to help us find a unified way of viewing the EPR correlations.

(Of course, if we explain EPR within the framework of a theory that introduces a physical asymmetry between the two wings, then Explanatory Symmetry allows causal asymmetry. For instance, "absolute-time" theories, in which one measurement precedes the other, may incorporate causal asymmetry without violating Explanatory Symmetry.)

In summary: Explanatory Symmetry is not motivated by specifically minimalist, generative, or continuity-causal intuitions. This constraint follows from the more general requirement that an explanation combine metaphysical constructs (such as causal connections) with physical description to provide a coherent way of looking at the events in question.

---

[49]Strictly speaking, an A-wing↔B-wing interchange "switches" an $\underline{\varepsilon}_a$=up, $\underline{\varepsilon}_b$=down joint measurement outcome into $\underline{\varepsilon}_a$=down, $\underline{\varepsilon}_b$=up, which is a different physical state of affairs. We can easily escape this difficulty, in two ways. One, we could consider two identical bosons in their triplet state. Those particles always yield $\underline{\varepsilon}_a$=up and $\underline{\varepsilon}_b$=up, or $\underline{\varepsilon}_a$=down and $\underline{\varepsilon}_b$=down. Alternatively, we could stick with our spin-1/2 particles in their singlet state, but build into our wing-interchange a spatial rotation that changes up into down and vice versa. Since rotational symmetry is fundamental in relativistic quantum theory, Explanatory Symmetry still holds.

4.4. Summary. The following table summarizes section 4 by briefly stating the status of Reich, Causal Unidirectionality, and Explanatory Symmetry with respect to the three notions of causation addressed in this paper.

	Reich	**Causal Unidirectionality**	**Explanatory Symmetry**
Minimalist causation	Holds, unless we renounce the intuition that each *independent* lawlike dependence correspond to a separate causal relation	Can fail, especially in spacelike case, since a symmetric lawlike dependence can correspond to a symmetric causal connection	Holds, since a causal explanation should mirror symmetries inherent in the complete physical description
Generative causation	Holds, because each independent lawlike correlation must be "brought about"	Holds, because the "generation" relation is intrinsically asymmetric	Holds, since a causal explanation should mirror symmetries inherent in the complete physical description
Continuity causation	Holds, except for non-Markovian processes, which cannot correspond to fundamental physical processes	Fails, unless we require (for instance) that a causal process correspond to physical transfer of a conserved quantity	Holds, since a causal explanation should mirror symmetries inherent in the complete physical description

5. CAUSAL NO-GO THEOREM

I now prove that any causal explanation of the EPR correlations consistent with Reich, Causal Unidirectionality, and Explanatory Symmetry is incompatible with the symmetric physical description provided by relativistic quantum theory. Then, I'll use the conclusions of section 4 (summarized in the table) to explore the philosophical implications.

Causal No-go Theorem:

Within the framework of relativistic quantum theory, we cannot causally explain the EPR correlations consistent with Reich, Causal Unidirectionality, and Explanatory Symmetry.

Proof:

Suppose that $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$ are directly causally connected. Then Causal Unidirectionality requires that exactly one of those two events be a (partial) cause of the other. For concreteness, let's say $\underline{\varepsilon}_a$ is a partial cause of $\underline{\varepsilon}_b$. Then Causal Unidirectionality requires us *not* to call $\underline{\varepsilon}_b$ a partial cause of $\underline{\varepsilon}_a$; but EPR Explanatory Symmetry requires us to call $\underline{\varepsilon}_b$ a partial cause of $\underline{\varepsilon}_a$. This contradiction prevents us from claiming that the two measurement outcomes are directly causally connected.

Therefore, we must claim that Ψ, or perhaps Ψ supplemented by other events, is the common cause of $\underline{\varepsilon}_a$ and $\underline{\varepsilon}_b$. But according to relativistic quantum theory,

$$p(\underline{\varepsilon}_b \mid \Psi, \underline{\varepsilon}_a) \neq p(\underline{\varepsilon}_b \mid \Psi).$$

Specifically, $p(\underline{\varepsilon}_b \mid \Psi)$ equals 1/2, while $p(\underline{\varepsilon}_b \mid \Psi, \underline{\varepsilon}_a)$ equals 0 or 1. This is true even if we build into Ψ the complete pre-measurement state of the universe. (According to quantum theory, the probability of $\underline{\varepsilon}_b$ depends on nothing other than $\underline{\varepsilon}_a$ and the pre-measurement quantum state of the particles.) Therefore, within the framework of relativistic quantum theory, Reich rules out any causal explanation of $\underline{\varepsilon}_b$ that doesn't include $\underline{\varepsilon}_a$ as a partial cause. But, as shown above, Causal Unidirectionality and EPR Explanatory Symmetry imply that a causal explanation of $\underline{\varepsilon}_b$ must *not* include $\underline{\varepsilon}_a$ as a partial cause. So, within the framework of relativistic quantum theory, one cannot

causally explain the EPR correlations consistent with Reich, EPR Explanatory Symmetry, and Causal Unidirectionality. *Q.E.D.*

6. IMPLICATIONS FOR EXPLAINING EPR CAUSALLY

I'll now spell out the philosophical implications of this no-go theorem.

6.1. Should we explain the EPR correlations generative-causally or minimalist-causally? A generative causation advocate cannot renounce Causal Unidirectionality or Reich (or Explanatory Symmetry). *Therefore, if she accepts relativistic quantum theory as fundamental, she must admit that some correlations in nature cannot be causally explained.*

Many previous articles, including Elby (1992), reach the italicized conclusion. Section 4 clarifies this conclusion by showing that it applies only to a specific notion of causation, namely generative causation. We *can* explain EPR within the context of minimalist causation or continuity causation.

Since minimalist causation is not committed to Causal Unidirectionality (for spacelike causal connections), a causal minimalist can explain the EPR correlations as follows: Ψ and ε_a are partial causes of ε_b, while Ψ and ε_b are partial causes of ε_a, end of story.[50] This causal story, however, seems not to have *explained* anything. Rather, this "explanation" merely calls "causal" the lawlike correlations encoded by relativistic quantum theory, without providing a deeper understanding of what's happening. A minimalist causation advocate could respond that we shouldn't *expect* anything more

[50]Alternatively, at great intuitive cost (see section 4.1), a causal minimalist can renounce Reich, and explain the measurement results in terms of a "non-screening-off common cause." Cartwright (1989) makes essentially this move.

from an explanation, because all talk of "bringing about" or continuous processes is nonsense, a pleasant way of helping us organize our thoughts. If you accept this response, however, then all lawlike dependencies in nature simply don't have a deeper explanation, and therefore you have little reason to assign causal relations to events. You might as well just catalog the lawlike dependencies and call it quits. Indeed, critics of minimalist causation often focus on its apparent explanatory emptiness.

Let me raise a brief sociological point: After examining the metaphysical pitfalls of generative causation and continuity causation, a philosopher might be tempted to embrace minimalist causation. I contend, however, that if an minimalist-causal explanation sounds appealing, it's only because the listener secretly fleshes out the explanation with generative-causal or continuity-causal intuitions. For example, an minimalist-causal explanation of why my TV turns on when I hit the remote control button would list the chain of lawlike dependencies between button pushings, cathode-ray tubes becoming warm, and so on. This catalog of dependencies appeals to our intuitions *because* we're secretly picturing infrared rays racing from the remote control to the TV, etc. These illicit continuity-causal images spice up the bare-bones minimalist-causal explanation to make it palatable. The minimalist-causal explanation, when stripped of these continuity-causal intuitions, seems just as non-explanatory as the minimalist-causal explanation of EPR. Nonlocal quantum correlations help us to entertain this criticism of minimalist causation, by providing a scenario in which we're less inclined to flesh out our minimalist-causal explanation with illicit continuity-causal or generative-causal intuitions.

6.2. Should we explain the EPR correlations continuity-causally?

A continuity causation advocate can explain EPR by renouncing Reich or Causal Unidirectionality. As long as "causal processes" mirror physical processes, Reich holds, because fundamental physical processes are Markovian. But we can renounce Causal Unidirectionality by allowing "passive" processes (i.e., those not corresponding to transfer of a conserved quantity) to count as causal. Healey's explanation of EPR violates Causal Unidirectionality, but obeys Reich. As discussed in section 4.3, such continuity-causal explanations are stripped of almost all generative-causal intuition.

These "eviscerated" continuity-causal explanations clarify issues both in philosophy of quantum theory and in philosophy of causation. The quantum philosopher, if he wants to explain the EPR correlations causally within the context of relativistic quantum theory, and if he considers minimalist causation to be explanatorily empty, must adopt a continuity-causal explanation in which the continuous causal process does not correspond to the flow of current density, energy, momentum, or any such quantity.

Should we label such a process "causal"? As discussed in subsection 3.4, in a classical framework, most continuity-causal explanations incorporate generative-causal intuitions, and vice versa. The intuitions urging you to call Healey's explanation "noncausal" are precisely those generative-causal notions normally present in continuity causation. On the other hand, any sympathies you feel for calling Healey's explanation "causal" stem from pure continuity-causal intuitions. So, EPR helps us to see what bare continuity causation looks like.

These considerations do not tell us what causation *really* is, or whether that question even makes sense. They merely help to clarify our options.

7. CONCLUSION

Redhead (1992), Elby (1992), and others attempt to show that we cannot causally explain the EPR correlations within a quantum framework. These arguments involve setting necessary conditions on causation. As Healey (1992) points out, however, causation is not a sufficiently univocal concept to invite universal necessary conditions. Therefore, the philosophical implications of causation no-go theorems are unclear.

In this paper, I tried to sharpen these no-go theorems by teasing apart three overlapping yet distinct notions of causation, namely minimalist causation, generative causation, and continuity causation. By exploring which necessary conditions each conception must adopt, we (at least partially) disentangled these causal notions. My no-go theorem showed that within a relativistic quantum framework, we cannot causally explain the EPR correlations consistent with Reich, Causal Unidirectionality, and Explanatory Symmetry. Therefore, a "causal" explanation of the EPR experiment within a relativistic quantum framework must be minimalist-causal or continuity-causal, stripped of almost all generative-causal intuition.

So, if you think all lawlike correlations warrant a causal explanation, you must severely water down your causal intuitions. Arguably, instead of watering down causation, we should search for a new, perhaps holistic framework in which to explain quantum correlations.

CHAPTER 6: CONCLUSION

Each chapter of this thesis was about a different topic. Chapter 2 explored nonlocality, and presented some new algebraic nonlocality proofs utilizing assumptions of unprecedented weakness. Chapter 3 argued that SQUID experiments say little about Macrorealism *per se*, but can rule out non-invasively measurability. In chapter 4, I showed how decoherence rescues modal interpretations from otherwise-fatal objections. And chapter 5 argued that any "causal" explanation of the EPR correlations will be severely watered down. In each chapter, I argued that the piece of "quantum weirdness" under discussion could be explained well by holism, the idea that composite systems possess properties that cannot even in principle be reduced to the properties of the parts. No one of my chapters presents a drop-dead argument that we should adopt a holistic explanatory framework. But taken together, my chapters point us in that direction.

REFERENCES

Albert, D. and B. Loewer (1988), "Interpreting the Many-Worlds Interpretation", *Synthese 77:* pp. 195-213.

Albert, D. and Loewer, B. (1990), "Wanted Dead or Alive: Two Attempts to Solve Schrödinger's Paradox", in A. Fine, M. Forces, and L. Weasels (eds.), *Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association, Volume 1.* East Lansing: Philosophy of Science Association, pp. 277-285.

Arntzenius, F. (1988), in *Proceedings of the 1988 Biennial Meeting of the Philosophy of Science Association, Volume 1.* East Lansing: Philosophy of Science Association.

Bacciagaluppi, G. and M. Hemmo (1995), "The Modal Interpretation of Imperfect Measurements", forthcoming in *Studies in the History and Philosophy of Modern Physics.*

Ballentine, L. (1987), *Physical Review Letters 59:* pp. 1493-

Bell, J. (1964), "On the Einstein-Podolsky-Rosen Paradox", *Physics 1:* pp. 195-200. Reprinted in Bell (1987), pp. 14-21.

Bell, J. (1966), "On the Problem of Hidden Variables in Quantum Mechanics", *Reviews of Modern Physics 38:* pp. 447-475. Reprinted in Bell (1987), pp. 1-13.

Bell, J. (1987), *Speakable and Unspeakable in Quantum Mechanics*. Cambridge: Cambridge University Press.

Bohm, D., B. Hiley, and P. Kaloyerou (1987), "An Ontological Basis for the Quantum Theory", *Physics Reports 144:* pp. 321-375.

Bransden, B. and C. Joachain (1989), *Introduction to Quantum Mechanics*. New York: Longman Scientific & Technical (co-published with Wiley).

Brown, H. and G. Svetlichny, "Nonlocality and Gleason's Lemma, part 1: Deterministic Theories", *Foundations of Physics 20:* pp. 1379-1387.

Cartwright, N. (1989), "Quantum Causes: The Lesson of the Bell Inequalities", in *Philosophy of the Natural Sciences: Proceedings of the 13th International Wittgenstein Symposium*. Vienna: Verlag Holder-Pichter Tempsky. pp. 120-127.

Cartwright, N., and M. Jones, M. (1991), "How to Hunt Quantum Causes", *Erkenntnis 35:* pp. 205-231.

Clauser, J. and M. Horne, "Experimental Consequences of Objective Local theories", *Physical Review D10:* pp. 526-535.

Clifton, R., M. Redhead, and J. Butterfield (1991), "Generalization of the Greenberger-Horne-Zeilinger Algebraic Proof of Nonlocality', *Foundations of Physics,* **21**, pp. 149-184.

Clifton, R. (1995), "Independently Motivating the Kochen-Dieks Modal Interpretation of Quantum Mechanics", forthcoming in *British Journal for the Philosophy of Science.*

Cohen-Tannoudji, C., B. Diu, and F. Laloë (1977), *Quantum Mechanic, volume 1.* Translated from the French by S. Hemley, N. Ostrowsky, and D Ostrowsky. New York: John Wiley & Sons.

D'Espagnat, B. (1976), *Conceptual Foundations of Quantum Mechanics.* Reading, Massachusetts: Addison-Wesley-Benjamin-Cummings.

deWitt, B. and R. Graham, (eds.), *The Many-Worlds Interpretation of Quantum Mechanics.* Princeton: Princeton University Press.

Dickson, M. (1994), "Wavefunction Tails in the Modal Interpretation", in D. Fine, M. Forces, and R. Burian (eds.), *Proceedings of the 1994Biennial Meeting of the Philosophy of Science Association, Volume 1.* East Lansing: Philosophy of Science Association. pp. 366-376.

Dieks, D. (1989), "Quantum Mechanics without the Projection Postulate and its Realistic Interpretation", *Foundations of Physics* 19: 1395-1423.

Dieks, D. (1994), "Objectification, Measurement and the Classical Limit According to the Modal Interpretation of Quantum Mechanics", in P. Busch, P. Lahti, and P.

Mittelstaedt (eds.), *Proceedings of the Symposium on the Foundations of Modern Physics, 1993*. Singapore: World Scientific.

Elby, A. (1990a), "On the physical interpretation of Heywood and Redhead's algebraic impossibility theorem", *Foundations of Physics Letters 3:* pp. 239-247.

Elby, A. (1990b), "Critique of Home and Sengupta's derivation of a Bell inequality", *Foundations of Physics Letters 3:* pp. 317-324.

A. Elby, (1990c) "Nonlocality and Gleason's lemma, part II: Stochastic theories", *Foundations of Physics 20:* pp. 1389-1397.

Elby, A. (1992), "Should we explain the EPR Correlations Causally?", *Philosophy of Science 59:* pp. 16-25.

Elby, A. (1993), "Why Modal Interpretations of quantum mechanics don't solve the measurement problem", *Foundations of Physics Letters 6:* pp. 5-19.

Elby, A. (1993), "Why local realistic theories must violate, nontrivially, the EPR-type perfect correlations", *British Journal for the Philosophy of Science 44:* pp. 213-230.

Elby, A. (1994), "Can decoherence solve the measurement problem?" in P. Mittelstaedt, P. Lahti, and P. Busch (eds.) *Symposium on the Foundations of Modern Physics 1993*. Singapore: World Scientific.

Elby, A. (1994), "The decoherence approach to the measurement problem", in D. Hull, M. Forces, and R. Burian (eds.) *Proceedings of the 1994 Biennial Meeting of the Philosophy of Science Association, volume 1*. East Lansing: Philosophy of Science Association. pp. 355-365.

Elby, A. and Sara Foster (1992), "Why SQUID experiments can rule out non-invasive measurability", *Physics Letters A 166:* pp. 17-23.

Elby, A. and M. Jones (1992) "Weakening the locality conditions in algebraic nonlocality proofs", *Physics Letters A 171:* pp. 11-16.

Elby, A., H. Brown, and S. Foster (1993), "What makes a theory physically complete?", *Foundations of Physics 23:* pp. 5-19.

Fine, A. (1974), "On the Completeness of Quantum Theory", *Synthese 29:* pp. 257-289.

Fine, A. (1982), "Hidden Variables, Joint Probability, and the Bell Inequalities", *Physical Review Letters 48:* pp. 291-295.

Fine, A. (1988), book review, *Foundations of Physics Letters 1:* p. 91.

Fine, A. (1989), *Foundations of Physics 14:* pp. 453-

Fleming, G. (1965), *Physical Review B139:* p. 963.

Foster, S. and A. Elby (1991), "A SQUID no-go theorem without macrorealism: What SQUID's really tell us about nature", *Foundations of Physics 21:* pp. 773-785.

Gleason, A. (1957), "Measures on the Closed Subspaces of a Hilbert Space", *Journal of Mathematics and Mechanics 6:* pp. 885-893.

Greenberger, D., M. Horne, A. Shimony, and A. Zeilinger (1990), "Bell Theorem Without Inequalities", *American Journal of Physics 58:* pp. 1131-1143.

Hardy, L. (1993), "Nonlocality for 2 Particles Without Inequalities for Almost All Entangled States", *Physical Review Letters 71:* pp. 1665-1668.

Healey, R. A. (1989), *The Philosophy of Quantum Mechanics: An Interactive Interpretation.* Cambridge: Cambridge University Press.

Healey, R. (1992), "Chasing Quantum Causes: How Wild is the Goose?", *Philosophical Topics 20:* 181-204.

Healey, R. A. (1994), "Nonseparable Processes and Causal Explanation", *Studies in History and Philosophy of Science 25:* pp. 337-374.

Hegerfeldt, G. (1974), *Physical Review D10:* p. 3320.

Hegerfeldt, G. (1985), *Physical Review Letters 54:* p. 2395.

Heywood, P. and M. Redhead (1983), "Nonlocality and the Kochen-Specker Paradox", *Foundations of Physics 13:* pp. 481-499.

Home, D. and G. Sengupta (1984), *Physics Letters A102:* pp. 159-162.

Jarrett, J. (1984), "On the Physical Significance of the Locality Conditions in Bell Arguments", *Noûs 18:* pp. 569-580.

Joos, E. and Zeh, H. D. (1985), "The Emergence of Classical Properties Through Interaction with the Environment", *Zeitschrift fur Physik B 59:* pp. 223-243.

Kochen, S. and E. Specker (1967), "The Problem of Hidden Variables in Quantum Mechanics", *Journal of Mathematics and Mechanics 17:* pp. 59-87.

Leggett, A. (1986a), in G. Grinstein and G. Mezenko (eds.), *Directions in Condensed Matter Physics: Memorial Volume in Honor of Shang-keng Ma.* Singapore: World Scientific. pp. 185-

Leggett, A.(1986b), in J. de Boer, E. Dal, and O. Ulfbect (eds.), *The Lesson of Quantum Theory: Niels Bohr Centenary Symposium.* Amsterdam: Elsevier. pp. 35-

Leggett, A. and A. Garg (1985), *Physical Review Letters 55* : pp. 857-

Lepore, V. and F. Selleri (1990), "Do Performed Optical Tests Disprove Local Realism?", *Foundations of Physics Letters 3:* pp. 203-220.

Lewis, D. (1986), *Philosophical Papers, Vol. II.* Clarendon Press: Oxford.

Mermin, N. D. (1990), "Extreme Quantum Entanglement in a Superposition of Macroscopically Distinct States", *Physical Review Letters 65:* pp. 1838-1840.

Peres, A. (1990), "Incompatible Results of Quantum Measurements", *Physics Letters A151:* pp. 107-108.

Redhead, M. (1987), *Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics.* Oxford: Clarendon Press.

Redhead, M. (1992), "Propensities, Correlations, and Metaphysics", *Foundations of Physics 22:* 381-394.

Reichenbach, H. (1956), *The Direction of Time.* Edited by M. Reichenbach. Berkeley: University of California Press.

Ruijsenaars, S. (1981), *Annals of Physics (New York) 137:* p. 33

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

Stairs, A. (1983), *Philosophy of Science 50:* pp. 587-

Stapp, H. (1993), "Significance of an Experiment of the Greenberger-Horne-Zeilinger Kind", *Physical Review A47:* pp. 847-853.

Stapp, H. (1993), "Stapp's Algebraic Argument for Nonlocality--Reply", *Physical Review A49:* pp. 4257-4260.

Suppes, P. and M. Zanotti (1976), "On the Determinism of Hidden-Variable Theories with Strict Correlation and Conditional Statistical Independence of Observables", in P. Suppes (ed.), *Logic and Probability in Quantum Mechanics.* Dordrecht: Reidel.

Svetlichny, G., M. Redhead, H. Brown, and J. Butterfield (1988), "Do the Bell Inequalities Require the Existence of Joint Probabilities?", *Philosophy of Science 55:* pp. 387-401.

Tesche, C. (1990), "Can a Non-invasive Measurement of Magnetic Flux Be Performed with Superconducting Circuits?", *Physical Review Letters 64* : pp. 2358-2361.

van Fraassen, B. (1973), "Semantic Analysis of quantum Logic", in C. Hooker (ed.), *Contemporary Research in the Foundations and Philosophy of Quantum theory.* Dordrecht: Reidel. pp. 80-113.

van Fraassen, B. (1979), "Hidden Variables and the Modal Interpretation of Quantum Mechanics", *Synthese 42:* pp. 155-165.

van Fraassen, B. (1991), *Quantum Mechanics: An Empiricist View*. Oxford: Clarendon Press.

Wheedan, R. and A. Zygmund (1977), *Measure and Integral: And Introduction to Real Analysis*. New York: Marvel Dekker, Inc.

Zeh, H. (1993), "There are No Quantum Jumps, nor are there Particles", *Physics Letters A172:* pp. 189-192.

Zurek, W. (1993a), "Preferred States, Predictability, Classicality, and the Environment-Induced Decoherence", *Progress in Theoretical Physics 89:* pp. 281-312.

Zurek, W. (1993b), "Negotiating the Tricky Border Between Quantum and Classical: Zurek Replies", *Physics Today* 46 no. 4: pp. 84-90.