Erschienen in : Synthese ; 128 (2001), 3. - S. 343-379
http://dx.doi.org/10.1023/A:1011971808127

THOMAS MÜLLER and TOMASZ PLACEK

AGAINST A MINIMALIST READING OF BELL'S THEOREM: LESSONS FROM FINE

ABSTRACT. Since the validity of Bell's inequalities implies the existence of joint probabilities for non-commuting observables, there is no universal consensus as to what the violation of these inequalities signifies. While the majority view is that the violation teaches us an important lesson about the possibility of explanations, if not about metaphysical issues, there is also a minimalist position claiming that the violation is to be expected from simple facts about probability theory. This minimalist position is backed by theorems due to A. Fine and I. Pitowsky. Our paper shows that the minimalist position cannot be sustained. To this end, we give a formally rigorous interpretation of joint probabilities in the combined modal and spatiotemporal framework of 'stochastic outcomes in branching space-time' (SOBST) (Kowalski and Placek, 1999; Placek, 2000). We show in this framework that the claim that there can be no joint probabilities for non-commuting observables is incorrect. The lesson from Fine's theorem is not that Bell's inequalities will be violated anyhow, but that an adequate model for the Bell/Aspect experiment must not define global joint probabilities. Thus we investigate the class of stochastic hidden variable models, which prima facie do not define such joint probabilities. The reason why these models fail supports the majority view: Bell's inequalities are not just a mathematical artifact.

1. INTRODUCTION

The majority view of Bell's inequalities endorses the following two points: (i) The inequalities have been violated by the Bell/Aspect quantum correlation experiment.[1] (ii) This empirical fact tells us something important about the way the world is (or at least, about the way the world can be explained): a rather general class of models that one would like to posit as the mechanism 'behind' the observed correlations, namely the class of so-called factorizable local hidden variable models, is unable to account for the phenomena, which in turn is seen as indicating a failure of locality, the presence of holistic features in quantum phenomena, or even a conspiracy in nature.

The second ingredient of the majority view is challenged by what we call the "minimalist interpretation" of the violation of Bell's inequalities. This interpretation is backed by two mathematical theorems due to Arthur Fine[2] and Itamar Pitowsky,[3] which link the violation of Bell's inequalities to some rather simple facts about probability theory. The minimalist interpretation sees the violation of Bell's inequalities as reflecting the nature of quantum mechanical probabilities rather than indicating anything about locality or holism in quantum phenomena.[4] A recent statement of this view is (De Beare et al., 1999, p. 68):

> (...) it is not nonlocality which is implied by the violation of Bell's inequality but, rather, the non-existence of joint probability for noncommuting single-particle operators (...).

Fine proved that the surface probabilities of the Bell/Aspect experiment are committed to Bell's inequalities if and only if there exists a joint probability distribution for all four quantum observables involved, some of which do not commute. Now Fine claims that it is 'the very essence of quantum mechanics' to deny that such distributions exist.[5] Fine's theorem can thus be taken to support the view that a violation of Bell's inequalities was to be expected in the first place and does not constitute a major experimental result.

Pitowsky's mathematical result is that an eight-tuple of real numbers from the interval [0,1], such as the experimental data from the Bell/Aspect experiment, can be viewed as a set of four single and four joint probabilities defined on a single classical probability space if and only if the eight-tuple satisfies Bell's inequalities. Again, the probabilities in a quantum correlation experiment involving non-commuting observables are not defined on a single probability space, but rather on four different probability spaces, so that one should not expect the experimental results to be embeddable in a single probability space. This can again be taken to support the view that the violation of Bell's inequalities was in fact to be expected.[6]

The aim of our paper is to show that the minimalist interpretation of the violation of Bell's inequalities cannot be sustained, even though the mathematical theorems supposedly backing the minimalist interpretation are correct. Our argument will proceed in three steps. First, we will argue against the view that the non-commutativity of quantum observables alone precludes one from defining a joint probability distribution for these observables. The upshot of this will be the following reading of Fine's theorem: since Bell's inequalities are violated in the Bell/Aspect experiment, the (non-commuting) quantum observables involved cannot have a joint probability distribution. In the second step, we will use the framework of 'stochastic outcomes in branching space-time' (SOBST),[7] to draw a distinction (similar to that of (Svetlichny et al., 1988)) between empirical and purely mathematical joint probabilities. The SOBST models allow for a unified treatment of causal, spatiotemporal and modal relations and are particularly suited for a discussion of joint probabilities in quantum mechanics, since most classical probability spaces have a representation in models of SOBST.[8] We will show that introducing an empirical joint probability to a SOBST model amounts to giving a determinate-value hidden variable model. It is further seen that in many cases, there is no penalty for introducing a joint probability for non-commuting observables; however, the introduction of joint probabilities for all the observables involved in the Bell/Aspect setup makes the resulting model empirically inadequate. In the third step, having taken a lesson from Fine's theorem, we attempt to construct a SOBST factorizable stochastic model for the Bell/Aspect experiment that uses four small probability spaces rather than a single large probability space defining joint probabilities for all the four observables involved. The dramatic fact is that, given the usual physically motivated constraints that factorizable stochastic models are subject to, the small probability spaces can be pasted together, so that global joint probabilities are mathematically definable, which makes the model empirically inadequate.

The paper is organized as follows: Section 2 gives an outline of the SOBST framework. Section 3 quickly reviews the mathematics of joint probability spaces, gives a precise statement of the results by Fine and Pitowsky, and shows their equivalence. Section 4 comments on various arguments to the effect that there can be no joint probability distributions for noncommuting quantum observables. Section 5 shows how to interpret joint probabilities in the SOBST framework, and proves the link between the existence of joint probabilities and determinate-value hidden variables. Section 6 discusses a factorizable stochastic model for the Bell/Aspect experiment. Section 7 comments on the relations between determinate-value and stochastic models. Two appendices contain the SOBST definition of a common cause and technical material from section 6.

2. THE FRAMEWORK OF SOBST

The framework to be presented here is a recent extension of Belnap's *branching space-time*[9] that has been augmented to include outcomes of events.[10] The principal algebraic feature of the framework is that the family of outcomes of a given event is a Boolean algebra and thus lends itself naturally to the introduction of probabilities.[11]

2.1. *Definitions*

We start with a partial order $\mathbf{W} = \langle W; \leq \rangle$. The elements of the nonempty set W are interpreted as spatiotemporal points understood as concrete particulars. The relation $x \leq y$ is interpreted as 'x is in the backward light cone of y', or, 'x can causally influence y'. We define: $x < y$ iff $x \leq y$ and $x \neq y$. This relation is extended to subsets of W:

DEFINITION 1. (of precedence). For $E, F \subseteq W$, $x \in W$ (1) $E \prec x$ iff $\forall e \in E\ e < x$; (2) $E \prec F$ iff $\forall x \in F\ E \prec x$.

As $\mathbf{W}$ allows for branching, two points can be separated not only spatiotemporally, but also modally, by belonging to incompatible courses of events.

DEFINITION 2. (of compatibility). $x, y \in W$ are *upward compatible* iff there is a $z \in W$ with $z \geq x$ and $z \geq y$; otherwise, they are called *upward incompatible* or *orthogonal* (written $x \perp y$). Conforming to standard mathematical usage, sets composed entirely of upward compatible elements are called *upward directed.*

Some special subsets of W will be called 'histories'. Intuitively, a history is to represent a possible course of events.

DEFINITION 3. (of a history). A subset h of W is a *history* iff h is a maximal upward directed subset of W (i.e., for all upward directed $h' \subseteq W$ we have: $h' \supseteq h$ implies $h' = h$). The set of all histories is denoted by $\mathcal{H}$.

This definition leads to counterintuitive results, if, for instance, our spatiotemporal world comes to an end. (Placek, 2000) indicates how to overcome this by building $\langle W, \leq \rangle$ from spatiotemporal histories rather than carving histories out of $\langle W, \leq \rangle$.

In line with the idea of branching, histories may split.

DEFINITION 4. (of splitting points). For any two orthogonal points $x, y \in W$, we define the set of splitting points $C(x, y) \subseteq W$ by putting $z \in C(x, y)$ iff z is a maximal element in $\{z \in W : z \leq x \ \&\ z \leq y\}$. If x and y are not orthogonal, we put $C(x, y) = \emptyset$.

To ensure that for any pair of orthogonal points x and y, $C(x, y)$ is nonempty and that sets of splitting points behave "nicely", we assume the following two conditions:

(C1) For any $x, y, z \in W$, if $x \perp y$ and $z \leq x$, $z \leq y$, then there is some $t \in C(x, y)$ with $t \geq z$.

(C2) For any $x, y, z, t \in W$, if $x \geq z$ and $y \geq t$, then $C(x, y) \supseteq C(z, t)$.

The following notions are required to introduce outcomes of events in W:

DEFINITION 5. (of relative orthogonality). Elements x, y of W are *orthogonal relative to* E, written $x \perp_E y$, iff $E \prec x$, $E \prec y$ and $C(x, y) \cap E \neq \emptyset$.

DEFINITION 6. (of orthogonal complement). For $F \subseteq W$, the *orthogonal complement of F relative to E* is the set $F^{\perp_E}$ such that $x \in F^{\perp_E}$ iff $\forall\, y \in F\ x \perp_E y$.

DEFINITION 7. (of outcome). A subset F of W is an *outcome* of $E \subset W$ iff $F = F^{\perp_E \perp_E}$

This definition ensures that an outcome of E is preceded by E and is located as close as possible to E. What the outcomes of E look like crucially depends on whether and, if so, how many, histories split in E. Given the above definitions, the following holds:[12]

THEOREM 1. The family F_E of outcomes of $E \subset W$ forms a complete and atomic Boolean algebra

$$\mathfrak{B}_E = \langle F_E, \cap, \cup, \perp_E, \mathbf{1}_E, \mathbf{0}_E \rangle,$$

where $\cap$ and $\cup$ are the familiar set-theoretical operations, the unit element of the algebra $\mathbf{1}_E = \{x \in W : E \prec x\}$, and the zero element of the algebra $\mathbf{0}_E$ is the empty set.

An *event* is defined as a subset of a history that is bounded from above:

DEFINITION 7. (of events). $E \subset W$ is an *event* iff $E \neq \emptyset$ and $\exists\, x \in W\ E \prec x$.

Only specific subsets of W have non-trivial outcomes. The following lemma shows that our definition of event is sensible:[13]

LEMMA 1. $E \subset W$ is an event iff E has a non-empty outcome.

In what follows, we will frequently refer to *atomic* outcomes of events:

DEFINITION 9. (of atomic outcomes). e is an atomic outcome of $E \subset W$ iff (1) e is a non-empty outcome of E and (2) there is no non-empty outcome u of E such that $u \subset e$.

In reasoning about correlations, we will need the notion of space-like separation:

DEFINITION 10. (of space-like events). The set $\{E_1, E_2, E_3, \ldots, E_n\}$ of events is *space-like* if and only if

- $\bigcup_{i=1}^{n} E_i$ is an event;
- E_i does not overlap with any outcome of E_j, i.e., for all $i, j \leq n$

$$E_i \cap \mathbf{1}_{E_j} = \emptyset.$$

In order to introduce probability spaces, it suffices to equip the Boolean algebra $\mathfrak{B}_E$ associated with the family of outcomes of an event E with a normalized, countably additive measure $\mu_E : \mathfrak{B}_E \to [0, 1]$, i.e., a measure satisfying

$$\mu_E(\mathbf{1}_E) = 1 \quad \text{and for mutually disjoint} \quad e_i \in \mathfrak{B}_E :$$

$$\mu_E(\bigcup_{i=1}^{\infty} e_i) = \sum_{i=1}^{\infty} \mu_E(e_i).$$

Outcomes extend as far into the future as possible: for outcome e of E and $a \in e$, $a \leq b$ implies $b \in e$. The probabilities in a quantum correlation experiment, on the other hand, are obtained by counting clicks or flashes, i.e., temporally bounded entities. In order to link the SOBST framework to such experiments, we cannot simply identify the clicks with outcomes. Rather, we assume that each atomic outcome e of a measurement event E starts with a *result* r^e, such as a click or a flash, and that we may identify the probability of the result r^e (established experimentally) with the probability $\mu_E(e)$ of the corresponding outcome e (defined in a SOBST model).

Just as events in a SOBST model give rise to Boolean algebras and thus, to probability spaces, the converse also holds:[14]

THEOREM 2 (Representation Theorem). Given a probability space $\langle \mathfrak{B}, \mu \rangle$ with $\mathfrak{B}$ complete and atomic, we can construct a SOBST model that contains an event E such that $\langle \mathfrak{B}, \mu \rangle = \langle \mathfrak{B}_E, \mu_E \rangle$.[15]

Within the probability space $\langle \mathfrak{B}_E, \mu_E \rangle$ associated with an event E, we can define conditional probabilities: the probability that outcome x of E happened, given that outcome y of E happened, is defined as usual as

$$\mu_E(x \mid y) = \begin{cases} \mu_E(x \cap y)/\mu_E(y) & \text{iff } \mu_E(y) \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The introduction of probabilities to the model **W** has the effect of making the resulting model more fine-grained. Thus, a model of *stochastic outcomes in branching space-time* (SOBST) is a quadruple $\langle W, \leq, \mathcal{E}, \Upsilon \rangle$, with W a non-empty set, $\leq$— a partial ordering on W, $\mathcal{E}$ — a set of events in W, and Υ — the collection of probability measures for each event in $\mathcal{E}$. In the next subsection, we will introduce some constraints on the assignment Υ so that we can consistently talk about joint events.

2.2. *Constraints on SOBST Models*

According to Definition 2, two points x and y are called *upward compatible* iff there is a point z with $x \leq z$ and $y \leq z$; otherwise, x and y are called *upward incompatible*. These notions are extended to events: two events E and F are upward compatible iff there is a point $z \in W$ such that $E \prec z$ and $F \prec z$. The upward compatibility of two events depends solely on the SOBST model in which they are defined — we can have SOBST models with incompatible events E and F that correspond to measurements of *compatible* observables in the sense of quantum mechanics. In order to avoid confusion, we will always use 'upward (in)compatible' explicitly for the SOBST notions. For upward compatible events, the following lemma holds:[16]

LEMMA 2. Given two upward compatible events E and F, the set-theoretical join $E \cup F$ is also an event, and the outcomes of $E \cup F$ a related to the outcomes of E and F as follows:

1. If e is an outcome of E and f is an outcome of F, then their set-theoretical intersection $e \cap f$ is an outcome of $E \cup F$.
2. If a is an outcome of $E \cup F$, then there are outcomes e of E and f of F such that $a = e \cap f$.

Furthermore, if e and f are atomic outcomes of E and F, respectively, then $e \cap f$ is either empty or an atomic outcome of $E \cup F$.

For a number of important simple cases with just one relevant spatial dimension, we can draw 2-dimensional graphs of SOBST models. Our convention is that the x-axis indicates spatial as well as modal separation, whereas the y-axis shows the temporal dimension of the model. Lines indicate the relation $\prec$ of precedence between subsets of W. (You may imagine arrowheads pointing upwards, in the $+t$ direction, everywhere.) Figure 1 shows a SOBST graph for two upward compatible events with two outcomes each, which illustrates Lemma 2.

If a SOBST model includes joint events in $\mathcal{E}$, a plausible probability assignment should somehow reflect this fact, thus imposing restrictions on the probability assignment Υ. We will consider two such restrictions: (i) the Locality Requirement, (ii) Screening-off.

The Locality Requirement. Frequently, in modeling a situation such as a correlation experiment, a SOBST model will contain two upward compatible events E and F with finitely many outcomes each such that for any two non-empty outcomes e and f of E and F, resp., the intersection $e \cap f$ is non-empty. (Intuitively, this means that 'no history escapes'.) In this case, it is required that the probabilities of the joint event $E \cup F$ return the single probabilities as marginals:

(1) $$\mu_{E\cup F}(e \cap \mathbf{1}_F) = \mu_E(e), \qquad \mu_{E\cup F}(\mathbf{1}_E \cap f) = \mu_F(f).$$

This condition is not a mathematical truth, but it should seem to be entirely reasonable and intuitive (if it did not hold, there might be no consistent way to talk about single probabilities when joint probabilities are defined). In fact, the Locality Requirement has two sides. For surface models like the model of Figure 2 (Section 3.2 below), in which all the probabilities are empirically established, the Locality Requirement is simply a testable empirical statement – you can do the experiment with one apparatus switched off, obtain the single probabilities $\mu_E(e)$ and $\mu_F(f)$, and compare with the marginals. When it comes to hidden variable models like the model of Figure 3 (Section 6 below), the status of the Locality Requirement changes, as it is now imposed on the hidden structure. The justification for this is the idea that whatever hidden structure we may wish to introduce, we should still be able to talk consistently about single probabilities when joint probabilities are defined. This form of the Locality Requirement is in fact quite strong, since it embodies the concept of non-contextuality and allows us to derive the usual Locality constraints for the Bell/Aspect experiment as well as the 'No Conspiracy' constraint that rules out influences from the measurement apparatus to the particle source (cf. Section 6).

Screening-off. The idea of Screening-off was introduced by (Reichenbach, 1956) in his attempt to provide an interpretation of causality in an indeterministic framework. It embodies an ideal of causal explanation.

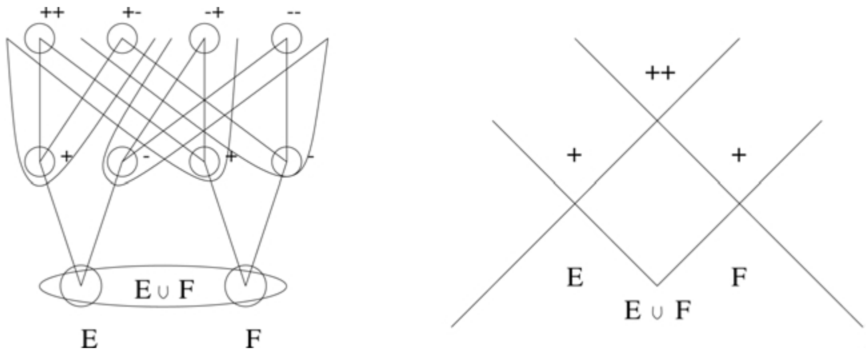

Figure 1. SOBST model with two upward compatible events E and F. Left: atomic outcomes '$+$' and '$-$' of E and F as well as the atomic outcomes '$++$', '$+-$', '$-+$', and '$--$' of the event $E \cup F$ are shown as a SOBST graph. Right: the relation between the atomic outcomes of the events E and F and the atomic outcomes of the event $E \cup F$ is illustrated for the case of joint outcome '$++$' via a space-time diagram: the outcome '$++$' of $E \cup F$ is the intersection of the outcome '$+$' of E and the outcome '$+$' of F.

The SOBST reading of Reichenbach's idea is that there should be no (unexplained) correlations between outcomes of space-like separated events (which by definition cannot causally influence one another), i.e., for space-like separated E and F and respective outcomes e and f, we should always have

$$(2) \qquad \mu_{E\cup F}(e \cap f) = \mu_E(e) \cdot \mu_F(f).$$

If, however, there are correlations between space-like separated events, i.e., if Equation (2) is violated for some e and f, then, according to Reichenbach, there has to be a common cause that explains these correlations: in an extended model with a common cause event C in the common past of E and F, Equation (2) holds for the outcomes of E and F conditional on the atomic outcomes ω_i of the common cause, i.e.,

$$(3) \qquad \mu_{E\cup F\cup C}(e \cap f \mid \omega_i) = \mu_{E\cup F\cup C}(e \mid \omega_i) \cdot \mu_{E\cup F\cup C}(f \mid \omega_i).$$ [17]

In the context of Bell's inequalities, the requirement that there be a common cause is the requirement that there be a certain hidden variable (h.v.) model. In Section 7 we will show that Screening-off is a much stronger criterion than one might expect.

3. THE MATHEMATICS OF JOINT PROBABILITIES AND FINE'S THEOREM

In this section, we will review some mathematical probability theory, state the theorems by Fine and Pitowsky mentioned in the introduction, and prove their equivalence.

In mathematics, one is accustomed to interpreting probability spaces as descriptions of chance experiments. Joint probability spaces thus have to be somehow connected to joint chance experiments. However, the interpretation of joint spaces, which is crucial in reasoning about quantum correlation experiments, is not straightforward. After comments on joint probabilities in QM in Section 4, Section 5 will therefore connect the mathematics with the SOBST framework, thus clarifying the issue.

In order to separate the mathematics clearly from the SOBST interpretation, we will use letters from the beginning of the alphabet for mathematical structures, whereas SOBST events will be denoted E, F etc. Also, we will use p for *purely mathematical probability measures* that have no SOBST interpretation, saving the letter μ for *empirical probability measures*, i.e., probabilities that have a SOBST interpretation in terms of weights on outcomes of events.

3.1. *The Mathematics of Joint Probabilities*

In sharp contrast to questions of interpretation, the mathematics of joint probability spaces is quite simple.

Given two Boolean algebras $\mathfrak{B}_A$ and $\mathfrak{B}_B$, their cartesian product, equipped with the obvious operations (e.g., $\langle a, b\rangle^\perp := \langle a^\perp, b^\perp\rangle$, $\langle a_1, b_1\rangle \wedge \langle a_2, b_2\rangle := \langle a_1 \wedge a_2, b_1 \wedge b_2\rangle$, $\mathbf{1}_{A,B} := \langle \mathbf{1}_A, \mathbf{1}_B\rangle$), is again a Boolean algebra. Thus, given two probability spaces $\mathcal{A} = \langle \mathfrak{B}_A, p_A\rangle$ and $\mathcal{B} = \langle \mathfrak{B}_B, p_B\rangle$, the Boolean algebra $\mathfrak{B}_{A,B} := \mathfrak{B}_A \times \mathfrak{B}_B$ will again admit of probability structures $\mathcal{A} \times \mathcal{B} = \langle \mathfrak{B}_{A,B}, p_{A,B}\rangle$, where the probability measure $p_{A,B}$ so far is entirely arbitrary. As a restriction on the probability measure $p_{A,B}$ that reflects the origin of the structure $\mathcal{A} \times \mathcal{B}$, it is natural to demand that $p_{A,B}$ return the single measures p_A and p_B as marginals, i.e.,

$$(4) \qquad p_{A,B}(\langle a, \mathbf{1}_B\rangle) = p_A(a), \qquad p_{A,B}(\langle \mathbf{1}_A, b\rangle) = p_B(b).$$

This restriction, called the *marginal property*, is easy to fulfill – in fact, there are generally infinitely many such measures. As an example, consider $\mathcal{A} = \mathcal{B}$, $\mathfrak{B}_A = \{\emptyset, a, b, \mathbf{1}\}$, $b = a^\perp$, $p(a) = p(b) = 1/2$. Then for any $r \in [0, 1/2]$, the measure $p_{A,B}$ with $p_{A,B}(\langle a, a\rangle) = p_{A,B}(\langle b, b\rangle) = r$,

$p_{A,B}(\langle a, b\rangle) = p_{A,B}(\langle b, a\rangle) = 1/2 - r$ will satisfy (4). A particularly simple joint measure is the *product measure*

(5) $\quad p_{A,B}(\langle a, b\rangle) := p_A(a) \cdot p_B(b).$

With the product measure, there will be no correlations of the form

(6) $\quad p_{A,B}(\langle a, b\rangle) \neq p_{A,B}(\langle a, \mathbf{1}_B\rangle) \cdot p_{A,B}(\langle \mathbf{1}_A, b\rangle) = p_A(a) \cdot p_B(b),$

whereas with other measures, such correlations will be present. – As the mathematical method outlined here builds one probability space from two, it can be iteratively extended to any finite number of probability spaces.[18]

These considerations show that mathematically, given any number of probability spaces, the existence of a joint probability space is always guaranteed. Furthermore, there is generally much freedom in defining a probability measure in the joint space. Thus, the mathematically interesting questions about joint probabilities are not of the form, "does a joint probability space for ... exist?", but of the form, "can one define a probability measure in the joint space so that it satisfies certain constraints?" We have seen that the constraint (4) can always be satisfied. It is only with more elaborate constraints that the question gets truly interesting.

3.2. *The Bell/Aspect Experiment, Random Variables, and Statistical Observables*

The Bell/Aspect quantum correlation experiment that we are considering has two wings with two possible settings in each, and each of the measurements has two possible outcomes (for the SOBST diagram, see Figure 2 below). This experiment has been chosen because it is one of the simplest setups that allow for a violation of Bell's inequalities and is thus rich enough for our aims. The model, and our argument, can easily be generalized to arbitrary 'finite' setups.

The restriction to finitely many outcomes allows us to keep to a simple notation in line with (Placek, 2000). The question about joint probabilities in QM is usually stated in terms of joint probability distributions for statistical observables, which in our framework we would have to define via random variables, and so far we have not even introduced random variables in the SOBST framework. While the introduction of random variables f_E in the SOBST framework poses no problem,[19] it is unnecessary in the finite case, and thus, we can stay at the level of probabilities of outcomes (or results). As we can take f_E to be injective, there is the natural correspondence $f_E(e) = r$, $\mu_E(f_E^{-1}(\{r\})) = \mu_E(e)$ for outcomes e of event E. In the Bell/Aspect case with two atomic outcomes '$+$' and '$-$'

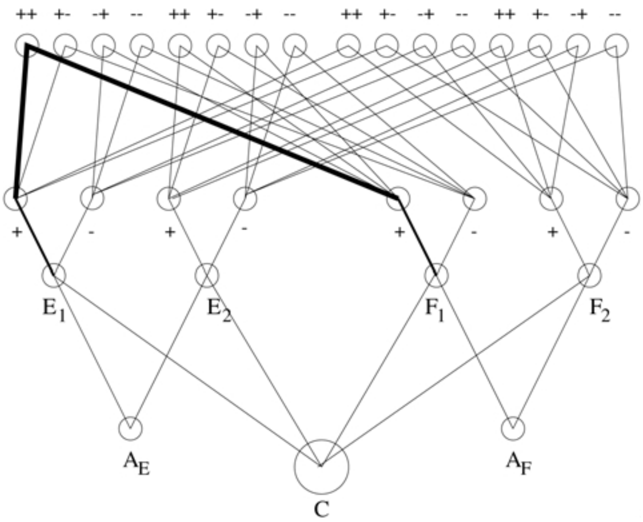

Figure 2. SOBST model of the Bell/Aspect experiment without hidden variables. C is the event of creating a particle pair, A_E (A_F) is the event of selecting the polarizer setting in the left (right) wing. E_i and F_j are the respective events of measuring the spin projection in the selected direction. Thick lines indicate a case of unexplained correlations: assuming, e.g., the polarizer setting in E_1 and F_1 to be parallel, we have $\mu_{E_1 \cup F_1}(+ \cap +) = 0 \neq \mu_{E_1}(+) \cdot \mu_{F_1}(+) = 1/2 \cdot 1/2$, even though events E_1 and F_1 are space-like separated.

and with the standard mapping $f_E(+) = 1$, $f_E(-) = -1$, we have accordingly $\mu_E(f_E^{-1}(\mathcal{R}^+)) = \mu_E(+)$ and $\mu_E(f_E^{-1}(\mathcal{R}^-)) = \mu_E(-)$ (where $\mathcal{R}^+ = \{x \mid x > 0\}$ and $\mathcal{R}^- = \{x \mid x < 0\}$).

3.3. *Fine's Theorem*

Fine's theorem establishes the equivalence of certain constraints on joint probability measures. Thus, on a purely mathematical reading, the theorem answers an 'interesting question' of exactly the form mentioned at the end of Section 3.1. In a watered-down version (which is sufficient for our purposes here), Fine's theorem states the following:[20]

THEOREM 3. Given four probability spaces $\mathcal{A}_i, \mathcal{B}_j, i, j = 1, 2$, each with two atoms '$+$' and '$-$', and four measures p_{A_i,B_j} in the joint probability spaces $\mathcal{A}_i \times \mathcal{B}_j$, $i, j = 1, 2$ that return the four probabilities p_{A_i} and p_{B_j} as marginals, the following two conditions are equivalent:

1. It is possible to define consistently a joint probability measure p_{A_1,A_2,B_1,B_2} on the Boolean algebra $\mathfrak{B}_{A_1} \times \mathfrak{B}_{A_2} \times \mathfrak{B}_{B_1} \times \mathfrak{B}_{B_2}$ that returns the four joint probabilities p_{A_i,B_j} (and thus *a fortiori* the four probabilities p_{A_i} and p_{B_j}) as marginals.

2. The eight given probability measures satisfy the following four (Bell/CH) inequalities

$$(7) \qquad -1 \leq p_{A_i,B_j}(+,+) + p_{A_i,B_{j'}}(+,+) + p_{A_{i'},B_{j'}}(+,+) - p_{A_{i'},B_j}(+,+) - p_{A_i}(+) - p_{B_{j'}}(+) \leq 0$$

   for $i, i', j, j' \leq 2, i \neq i', j \neq j'$.

3.4. *Pitowsky's Theorem*

Pitowsky's theorem is formally similar to Fine's theorem, since it also links the fulfillment of inequalities to the possibility of constructing a probability space: a given vector of numbers can be treated as probabilities defined on a single probability space if and only if the numbers satisfy certain inequality constraints.

THEOREM 4. Given an 8-tuple $s = \langle p_1, p_2, p_3, p_4, p_{13}, p_{14}, p_{23}, p_{24}\rangle$ of real numbers from the interval [0,1], the following three conditions are equivalent:

1. The 8-tuple s is an element of the Clauser-Horne correlation polytope $c(4, S)$ with $S = \{\{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}\}$, i.e., it can be represented as a weighed classical truth value assignment for four propositional variables $p_1, \ldots, p_4$ and their four conjuncts $p_1\&p_3, \ldots, p_2\&p_4$.[21]
2. There exists a probability space $\langle \mathfrak{B}, p\rangle$ and $A_1, A_2, A_3, A_4 \in \mathfrak{B}$ such that

$$(8) \qquad p_k = p(A_k) \text{ and } p_{ij} = p(A_i \cap A_j) \quad \text{for } k \leq 4, i = 1, 2, \ j = 3, 4.$$

3. The numbers p_i and p_{ij} satisfy the following system of inequalities:

$$(9) \qquad -1 \leq p_{ij} + p_{ij'} + p_{i'j'} - p_{i'j} - p_i - p_{j'} \leq 0,$$
$$p_i \geq p_{ij}, \quad p_j \geq p_{ij}, \quad p_i + p_j - p_{ij} \leq 1$$

   for $i, i' = 1, 2, i \neq i'$ and $j, j' = 3, 4, j \neq j'$.

3.5. *Fine's and Pitowsky's Theorems are Equivalent*

Pitowsky claims that his theorem is equivalent to Fine's.[22] Still, there are dissenting views.[23] After all, Fine's theorem is about eight given probability spaces, whereas Pitowsky's is about eight given real numbers, so the premises of the theorems are different. However, we can identify the probability $p_{A_i}(+)$ mentioned in Theorem 3 with the number p_1 from Theorem 4, $p_{A_1,B_1}(+,+)$ with p_{13}, etc., thus unifying the premises.

We first prove that Theorem 3 entails Theorem 4. Given the eight probability spaces $\mathcal{A}_i$, $\mathcal{B}_j$, and $\mathcal{A}_i \times \mathcal{B}_j$, $i, j = 1, 2$, as in the premises of Theorem 3 and satisfying clause 2 of the theorem, the 8-tuple $s = \langle p_{A_1}(+), p_{A_2}(+), p_{B_1}(+), p_{B_2}(+), p_{A_1,B_1}(+,+), \ldots, p_{A_2,B_2}(+,+)\rangle$ is an element of $[0, 1]^8$ satisfying clause 3 of Theorem 4.

Proof. $s \in [0, 1]^8$ is clear, since the elements of s are probabilities. As to the inequalities (9), the four of the first line are exactly inequalities (7). In the second line, $p_i \geq p_{ij}$ and $p_j \geq p_{ij}$ follow from the premises of Theorem 3, as p_i and p_j are marginals of p_{ij} (thus, e.g., $p_1 = p_{A_1}(+) = p_{A_1,B_1}(+,+) + p_{A_1,B_1}(+,-) \geq p_{A_1,B_1}(+,+) = p_{13}$). The last inequality follows since, e.g., $p_1 + p_3 - p_{13} = p_{A_1}(+) + p_{B_1}(+) - p_{A_1,B_1}(+,+) = 1 - p_{A_1,B_1}(-,-) \leq 1$. □

In the other direction, given an 8-tuple $s \in [0, 1]^8$ satisfying clause 3 of Theorem 4, we can construct uniquely eight probability spaces satisfying the premises and clause 2 of Theorem 3 by taking $p_{A_1}(+) = p_1$, $p_{A_1}(-) = 1 - p_1, \ldots, p_{B_2}(-) = 1 - p_4$, $p_{A_1,B_1}(+,+) = p_{13}$, $p_{A_1,B_1}(+,-) = p_1 - p_{13}$, $p_{A_1,B_1}(-,+) = p_3 - p_{13}$, $p_{A_1,B_1}(-,-) = 1 + p_{13} - p_1 - p_3$, $\ldots, p_{A_2,B_2}(-,-) = 1 + p_{24} - p_2 - p_4$.

Proof. The construction is unique. The fact that $s \in [0, 1]^8$ guarantees that the probability spaces $\mathcal{A}_i$ and $\mathcal{B}_j$, $i, j = 1, 2$, are well-defined. By $s \in [0, 1]^8$ and the second line of (9), the probability spaces $\mathcal{A}_i \times \mathcal{B}_j$, $i, j = 1, 2$, are also well-defined (i.e., the probabilities of their atoms $(+,+), \ldots, (-,-)$ are in $[0, 1]$ and sum up to 1). The inequality (7) holds because of the first line of (9). By construction, the measures p_{A_i,B_j} return the right marginals. □

Thus, Fine's and Pitowsky's theorems are equivalent. In what follows, we will always discuss Fine's theorem; the implications for Pitowsky's result should be obvious.[24]

4. THE ISSUE OF JOINT PROBABILITIES IN QUANTUM MECHANICS

A natural reading of Fine's theorem, and one suggested by Fine himself, shows that if the Bell/CH inequalities hold, then a joint probability distribution for all the observables in the Bell/Aspect experiment is definable. Now, some of these observables do not commute, and it is a well-entrenched view that QM prohibits the existence of joint probability distributions for non-commuting observables.[25] Thus, the minimalist interpretation, via *modus tollens*, argues that the Bell/CH inequalities must be violated on purely mathematical grounds, quite independently of any considerations about hidden variable models or locality.[26] In this section, we will comment on the rationales for accepting the premise that joint probability distributions for non-commuting observables are prohibited.

In classical mechanics, experimental propositions correspond to projectors on subsets of a system's phase space. The algebra of these projectors, which is the usual set algebra, is Boolean. Thus, it is unproblematic to introduce a global probability measure providing for joint probabilities for any experimental propositions whatsoever.

In standard quantum mechanics, the situation is different. Experimental propositions correspond to projectors on closed subspaces of a Hilbert space, and the ensuing algebra, which is not the usual set algebra, is non-Boolean. If attention is restricted to commuting observables, the resulting algebra is again Boolean and allows for joint probabilities. However, in the general case of non-commuting observables, joint probabilities are problematic.

In what follows, we will assess the claim that joint probabilities for non-commuting quantum observables are not just problematic, but in fact prohibited. We will consider four arguments of increasing strength that purport to show this.

4.1. *Non-Commuting Observables are Not Comeasurable*

It is commonly held that observables that QM classifies as non-commuting cannot be measured simultaneously. Is this premise already sufficient to argue that there can be no joint probabilities for non-commuting observables?

Let us assume that two measurements that cannot be carried out simultaneously are characterized by probability spaces $\langle \mathfrak{B}_1, p_1 \rangle$ and $\langle \mathfrak{B}_2, p_2 \rangle$, respectively. We are asking whether there exists a joint probability measure p_{12} on $\mathfrak{B}_1 \times \mathfrak{B}_2$ such that $p_{12}(\langle a_1, a_2 \rangle)$ can be interpreted as the probability that measurement 1 yields the result a_1 *and* measurement 2 yields the result a_2 – keeping in mind that the two measurements cannot be carried out simultaneously.

To put the matter down to earth, imagine that (contrary to the facts) no two fair Polish coins can be tossed simultaneously, and that $\langle \mathfrak{B}_1, p_1 \rangle$ and $\langle \mathfrak{B}_2, p_2 \rangle$ describe tosses of two such coins. By the product probability construction (eq. (5)), you arrive at a probability measure $p_{12} = p_1 \cdot p_2$. However, you cannot read, say, $p_{12}(\langle t_1, h_2 \rangle)$ as the probability of simultaneously obtaining tails on the first coin and heads on the second coin: assigning a positive value to a probability like $p_{12}(\langle t_1, h_2 \rangle)$ is at odds with the fact that the two coins cannot be tossed simultaneously. On the other hand, if you assign all probabilities p_{12} the value zero, you will not recover the marginals, e.g., $p_1(t_1) = p_2(h_2) = 1/2$. Is there any way out of this quandary? Yes, there are actually two ways out: (i) You may assign zero to all the probabilities p_{12}, but then you must consequently read the returned marginal probabilities, e.g., $p(h_1) = 0$, as the probability that *in this kind of experiment, i.e., with two coins tossed simultaneously*, the first coin lands heads up. (ii) You may also take any p_{12} that returns as marginals the single probabilities like $p_1(t_1) = p_2(h_2) = 1/2$, but then you must consequently refuse to interpret the probabilities p_{12} as *empirical*, i.e., you may not read them as the probability of simultaneously obtaining a result on the first coin together with a result on the other coin: the probabilities p_{12} will be purely mathematical probabilities.

The upshot of this discussion is that the impossibility of performing two measurements simultaneously does not speak against defining corresponding joint probabilities, although the probabilities may have a purely mathematical meaning.

4.2. *Joint Probabilities are not Defined in QM*

The second argument against joint probability distributions for non-commuting quantum observables cites a simple matter of fact: The probabilistic algorithm of QM does not define such distributions. The algorithm defines joint probabilities for measurement results of pairs of commuting observables, but it fails in the case of non-commuting observables. The algorithm says that if D is a quantum system's state, A and B are observables, e and f are Borel sets on real line, and $P_A(e)$ and $P_B(f)$ are the respective projectors from the spectral resolution of A and B, then we get the following probability:

$$(10) \qquad p^D_{A,B}(e \times f) = tr(D\, P_A(e)\, P_B(f)).$$

This expression makes sense only for commuting A and B, because only commuting observables resolve into projectors such that for all Borel sets e and f, $P_A(e)\, P_B(f)$ is again a projector, and this is a necessary condition for the right hand side of equation (10) to be meaningful. For commuting A and B, the formula naturally gives the probability that, for a system in state D, the measurement of A yields a result in e and the measurement of B yields a result in f; for non-commuting A and B, the formula is meaningless.

However, the fact that the probabilistic algorithm of standard QM does not define joint probabilities for non-commuting observables does not speak against defining them otherwise. As QM does not probably have the last word about physical reality, one may still believe that a next theory will ascribe joint distributions to non-commuting observables of QM.

4.3. *Joint Probabilities Cannot be Defined*

The third argument draws on the most abstract layer of standard QM – its algebraic structure. As was mentioned above, experimental propositions like "the measurement of A yields a result from e" correspond to projectors on closed subspaces of an appropriate Hilbert space $\mathfrak{H}$. Thus, the family of experimental propositions about a QM system is characterized by the set of closed subspaces of $\mathfrak{H}$, which algebraically has the structure of an orthomodular lattice – a structure that is weaker than the Boolean algebra structure of classical experimental propositions, on which classical (Kolmogorov) probability spaces can be defined. Thus, the problem is to ascribe probabilities to an orthomodular lattice. This can be affected by structures usually called "generalized probability spaces". These probability spaces should be used to define joint probabilities for quantum observables. Now, classical joint probabilities satisfy certain natural properties, e.g., the marginal property (cf. Equation (4) above). The dramatic mathematical fact is that, generally speaking, not all these properties can be supported by probabilities defined on an orthomodular lattice. However, if only commuting observables are considered, joint probabilities with the relevant properties can be defined.[27] The immediate moral of this fact is that given the algebraic structure of standard QM, joint probabilities for non-commuting observables *cannot* be defined (in addition to the simple "are not defined" of the last argument) . Moreover, in order to define them in a generalized theory, there is a a rather high price to be paid: roughly, the algebraic structure of this theory must be "stronger" than an orthomodular lattice.

However, if one wants to build on these facts a general argument to the effect "no joint probability distributions for non-commuting observables", this will suffer the same drawback as the analogical move in the last argument: QM probably does not have the last word about the physical world, and a superseding theory may solve the matter differently.

4.4. *Joint Probabilities Cause Trouble*

The data of the Bell/Aspect experiment violate the Bell/CH inequalities. Via Fine's theorem, this means that in this experiment, the existence of joint probabilities for all the four quantum observables involved, two pairs of which do not commute, causes empirical trouble. This observation could easily be generalized to the following claim: there is an empirical penalty for defining joint probabilities for non-commuting quantum observables.

A number of theorems seem to point in this direction,[28] but the claim is in fact false.[29]

Note that already in the discussion of Section 4.1., we defined mathematically a joint probability for outcomes of two incompatible measurements, e.g., measurements of non-commuting quantum observables. This is already sufficient to refute the above claim. In Section 5 we will use the SOBST framework to shed more light on this issue, explicitly giving a reading of joint probabilities in terms of hidden variables.

4.5. *Assessment of the Arguments*

We have analyzed four arguments against joint probabilities for non-commuting quantum observables. The first and the fourth argument were seen to be fallacious, whereas the second and third established that in standard QM, such joint probabilities are not defined, and cannot be defined either. What these arguments fail to establish is the general claim that any two observables that standard QM characterizes as non-commuting cannot have joint probability distributions in some future theory. This claim can only be supported by a leap of faith committing one to the truth of an abstract layer of QM. In terms of the data, it is not warranted.

We started this section by pointing out that the minimalist interpretation would like to use Fine's theorem to support the following argument: (i) joint probabilities for non-commuting quantum observables cannot be defined, (ii) in the Bell/Aspect experiment, non-commuting quantum observables are involved. Therefore, (iii) no joint probabilities can be defined for the observables in the experiment. Via Fine's theorem, this means that (iv) the Bell/CH inequalities will be violated. In this argument, neither hidden variables nor locality considerations etc. are mentioned; thus (v) the violation of the inequalities is a purely mathematical fact.

We have now seen that the known arguments purportedly establishing (i) are a failure. Furthermore, (i) is not just unsupported, but actually false. Thus, the attempt of using Fine's theorem to support the minimalist interpretation is a non-starter. Still, there is a lesson to be learned from the theorem, if you read it the other way round: As the Bell/CH inequalities are violated in the Bell/Aspect experiment, joint probabilities for all the four observables involved must not be definable in an adequate model for the experiment. Fine's theorem should thus guide your search for an adequate model. We will come back to this in Section 6.

5. JOINT PROBABILITIES IN SOBST MODELS

As we already announced in Section 4.4,, we will now show (i) that it is not the case that introducing joint probabilities for non-commuting observables will generally lead to empirically inadequate models and (ii) how to interpret such joint probabilities in the SOBST framework. To this end, we will first address the SOBST interpretation of joint probabilities in a very general way. Recall the distinction made in Section 3: With respect to a given SOBST model, we call a mathematical probability space $\langle \mathfrak{B}, p \rangle$ *empirical* iff it is the probability space $\langle \mathfrak{B}_E, \mu_E \rangle$ of the outcomes of an event E in the given model; otherwise, we call the probability space $\langle \mathfrak{B}, p \rangle$ *purely mathematical*.

Assume that we are given a SOBST model with two events E and F and associated probability spaces $\mathcal{E} = \langle \mathfrak{B}_E, \mu_E \rangle$ and $\mathcal{F} = \langle \mathfrak{B}_F, \mu_F \rangle$, respectively. We have seen in section 3.1 that mathematically, the existence of a joint probability space $\mathcal{E} \times \mathcal{F} = \langle \mathfrak{B}_{E,F}, p_{E,F} \rangle$ is always guaranteed, and that we may require the joint probability measure $p_{E,F}$ to have the marginal property (eq. (4)). The question we are going to consider now is, can we give a SOBST interpretation of the joint space $\mathcal{E} \times \mathcal{F}$ such that the initially purely mathematical probability $p_{E,F}$ will be empirical?

5.1. *The Joint Space $\mathcal{E} \times \mathcal{F}$*

From the representation Theorem 2 we know that whenever there is a probability space $\mathcal{A} = \langle \mathfrak{B}_A, p_A \rangle$ with $\mathfrak{B}$ complete and atomic (conditions that are always satisfied in the finite cases considered here), there exists a SOBST model with an event E the outcomes of which have the Boolean algebra structure of $\mathfrak{B}_A$ and the respective probability measure p_A. This means that for the space $\mathcal{E} \times \mathcal{F}$, there is a SOBST model with an event C the outcomes of which have the structure of $\mathfrak{B}_{E,F}$ and the probabilities $p_{E,F}$. However, nothing guarantees that the initial SOBST model in which E and F are defined has any connection with the model in which C is defined. So far, Theorem 2 is of little help – with respect to the initial model, $p_{E,F}$ is still purely mathematical. We will now show how to construct a unified model for the events E, F, and C, i.e., a model in which $p_{E,F}$ is an empirical probability measure. To this end, we will need to distinguish carefully the case when E and F are upward compatible from the case of upward incompatible events.

Upward compatible events. In Section 2.2 we have seen that for upward compatible events E and F of the kind appropriate for modeling a correlation experiment (i.e., such that for non-empty outcomes e of E and f of

F, $e \cap f$ is non-empty), SOBST with the Locality Requirement gives a natural interpretation of the joint space:

$$(11) \qquad \mathfrak{B}_{E,F} = \mathfrak{B}_{E \cup F}, \quad p_{E,F}(\langle e, f \rangle) = \mu_{E \cup F}(e \cap f).$$

Thus, the event C mentioned above can in this case be identified with the event $E \cup F$. This case is completely unproblematic, both from the point of view of SOBST and from the point of view of QM – upward compatible events can only model comeasurable, i.e., commuting, quantum observables. Moreover, it is this construction that we need in familiar applications of joint probabilities in describing multiple chance experiments, like multiple coin tosses. In the case of two coins, e.g., the joint probability $p_{E,F}(r_1, r_2)$ of obtaining result r_1 on the first coin and result r_2 on the second coin can be interpreted as the probability $\mu_{E \cup F}(e_1 \cap e_2)$ of the outcome $e_1 \cap e_2$ of the event $E \cup F$ of tossing the two coins, where e_i is the atomic outcome of E (the event of tossing the first coin) or F (the event of tossing the second coin) corresponding to result r_i.

Upward incompatible events. Assume that $\mathcal{E}$ and $\mathcal{F}$ are probability spaces associated with upward incompatible events E and F in a SOBST model, such as measurements of non-commuting quantum observables, and that we are given a purely mathematical probability measure $p_{E,F}$ that returns the measures μ_E and μ_F as marginals (recall from Section 3.1 that the existence of such a measure is always guaranteed).

As E and F are upward incompatible, $E \cup F$ is not an event, and we need to read the probability measure $p_{E,F}$ as the probability for outcomes of an event C *different* from $E \cup F$. If this event C is to be incorporated in the given SOBST model, the only possible choice for the location of C is the common past of E and F, i.e., $C \prec E$ and $C \prec F$ – for C to be an event, it has to be a segment of one history, and within or after E and F, histories have split already. We will require for simplicity that all atomic outcomes of C contain E or F. (It will be helpful to picture a situation where E and F occur after C in two alternative futures of C.) Then, $E \cup C$ and $F \cup C$ will be events as well, with their outcomes satisfying (via the Locality Requirement) for all outcomes e, f, and ω of E, F and C, resp.:

$$(12) \qquad \mu_{E \cup C}(e \cap \mathbf{1}_C) = \mu_E(e), \qquad \mu_{E \cup C}(\mathbf{1}_E \cap \omega) = \mu_C(\omega),$$

$$\mu_{F \cup C}(f \cap \mathbf{1}_C) = \mu_F(f), \quad \mu_{F \cup C}(\mathbf{1}_F \cap \omega) = \mu_C(\omega).$$

It remains to give a SOBST interpretation for $p_{E,F}$.

5.2. *Upward Incompatible Events: The Joint Measure $p_{E,F}$*

With E and F upward incompatible and C in their common past, histories leading to E or F will split somewhere. This could be in C or not. If the splitting occurs in C, C will be an event of selecting between E and F. The more general option is to let the splitting depend on influences from an event A outside of C (think of the independent agent or the random device selecting among the possible directions of spin projection measurement in the Bell/Aspect experiment). C has thus no power over which of the upward incompatible events actually occurs, but the outcomes of E and F will be differentiated through the outcomes of C.

The mathematically given joint measure $p_{E,F}$ is defined on pairs $\langle e, f\rangle$, with e an outcome of E and f an outcome of F. The outcomes of C must correspond one-to-one to such pairs, and the probability for each of the outcomes of C must be the probability for the corresponding pair. Thus, to take a simple example, if E and F both have two atomic outcomes '$+$' and '$-$', C has four atomic outcomes $\omega_{\pm,\pm}$, which are assigned probabilities $\mu_C(\omega_{x,y}) = p_{E,F}(\langle x, y\rangle)$. The requirement that $p_{E,F}$ return the given single probabilities as marginals yields

$$\mu_E(x) = \mu_C(\omega_{x,+}) + \mu_C(\omega_{x,-}), \tag{13}$$

$$\mu_F(y) = \mu_C(\omega_{+,y}) + \mu_C(\omega_{-,y}).$$

These equations suggest reading μ_E and μ_F as *surface probabilities* that can be derived from underlying *hidden probabilities* μ_C. The move of incorporating the event C into the initial model allowed us to read the purely mathematical joint probabilities $p_{E,F}$ as empirical probabilities μ_C, namely, as the probability of outcomes of the event C. On this reading, the atomic outcomes $\omega_{\pm,\pm}$ of C differentiate the outcomes of E and F into two kinds: the outcome '$+$' of E, e.g., can happen as one of $\omega_{+,+}$ and $\omega_{+,-}$ (cf. the left part of Figure 3 for an illustration).

This situation allows one to interpret the atomic outcomes $\omega_{\pm,\pm}$ of C as instruction sets for both E and F, i.e., as *determinate-value hidden variables*: If, e.g., C results in $\omega_{-,+}$, then if E occurs, the outcome will be '$-$', whereas if F occurs, the outcome will be '$+$' (cf. the middle and right part of Figure 3). On this interpretation, the occurrence of E or F (which is governed by event A) has the effect of reading out the first or the second slot of the instruction set provided by the outcome of C. The model is subject to *counterfactual definiteness*: the hidden state $\omega_{x,y}$ predetermines not only the outcome of the event that actually occurs, but also of the event that does not occur and about which we can only speak counterfactually (the event that could have occurred). Thus, there is no genuine splitting (no

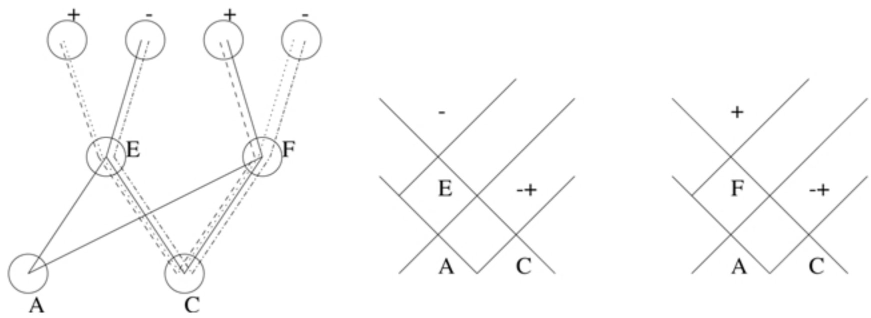

Figure 3. Determinate-value ('fatalistic') hidden variable model for two upward incompatible events E and F. Left: A SOBST graph shows the causal relations between the event A of selecting either E or F, the ('hidden') event C providing the instruction sets (indicated by four different line types), the events E and F, and the surface outcomes '$+$' and '$-$' of E and F. As the graph shows, each surface outcome in the hidden variable model integrates two 'hidden' outcomes of C. E.g., the outcome '$+$' of F comprises outcomes ω_{++} (dashed) and ω_{-+} (solid) of C. Middle and right: two space-time diagrams for the outcome ω_{-+} of C illustrate the epistemic status of the splitting at A, justifying the name 'fatalistic model'. In the middle, A selects E, while C has already given the instructions ω_{-+}. Accordingly, E reads out the outcome '$-$'. On the right, A selects F, resulting in outcome '$+$'.

objective chance) in the events E or F any more – the splitting has already occurred in C. (This does not mean that there can be no objective chance in the model any more: A as well as C may still operate indeterministically.) From Equation (13) it follows that on this reading, the surface probabilities will be correctly reproduced. Even though A and C may still host objective chance, we call determinate-value models *fatalistic*: from the point of view of E and F, the outcomes are selected on a pure 'it was to be' basis.

This brings us to the modal aspect of the situation. Initially, by specifying in SOBST two upward incompatible events E and F, there were two successive levels of modality present: (i) which of the two events actually takes place?, and (ii) which outcome will the actually occuring event have? These modalities corresponded (i) to splittings in A and (ii) to splittings in E *or* in F, which occurred after A. In the hidden variable model, the splittings in E or F have been cut short, as C already predetermines the outcomes of both E and F. Thus, there are no two successive levels of modality in the model any more: the splitting in A, while still present on the surface description (so that the model can be adequate), does not 'set the stage' for a further splitting in E or F, but only determines which of the two slots of the instruction sets given through the outcomes of C is read out. If we know which outcome of C we are in, we already know the results of E and F, no matter which of the two actually occurs. In the hidden variable model, the splitting in A thus has a purely epistemic status

– even if we might be limited in principle to reading out only one slot. Therefore, the hidden variable model has affected a collapse of modalities to just one genuine level, which is provided by the splittings in C.

Nothing in this argument depends on the fact that E and F have two atomic outcomes each, like in the example just given. In fact, if E and F have any finite number n_E and n_F of atomic outcomes, resp., then taking C to have $n_E \cdot n_F$ atomic outcomes will allow for exactly the same fatalistic construction. More generally, for any number of events with finitely many outcomes each, the mathematical existence of global joint probabilities allows the above construction of a fatalistic hidden variable model. In the Bell/Aspect experiment, this means that if there is a joint probability measure p_{E_1,E_2,F_1,F_2} for all the four observables involved, then one can construct a determinate-value hidden variable SOBST model for the experiment.

5.3. *Empirical Implications*

There is nothing wrong empirically with the mathematical existence of a joint probability measure for two upward incompatible events E and F, nor with the possibility of giving a determinate-value hidden variable model, even if E and F are taken to represent incompatible quantum measurement events (i.e., measurements of non-commuting observables of a quantum system). Standard 'quantum-mechanical wisdom' will only dictate that in this case E and F are not comeasurable, and accordingly, that $E \cup F$ is not an event, which is adequately captured in the SOBST model for the situation given here that takes E and F to be upward incompatible. Empirically, any assignment of probabilities $p_{E,F} = \mu_C$ for the hidden states, i.e., any probability measure on the mathematical joint space, will yield an adequate picture as long as the measure returns μ_E and μ_F as marginals.

Note that already by defining the joint probability $p_{E,F}$, we have left the orthodox quantum mechanical framework: in quantum mechanics, the joint probability $p_{E,F}$ that could be introduced purely mathematically is not defined for incompatible measurements.[30] The interesting fact is that while we may well have violated the spirit of QM through our construction, no empirical penalty is forthcoming so far. Thus, we have the following result: *It is always possible to introduce a determinate-value hidden variable model for the outcomes of two incompatible events with finitely many outcomes* – for two events, fatalism is always an option.

To sum up this section, we have cashed in our announcement of Section 4.4: *contra* the minimalist interpretation, there may exist joint probabilities for non-commuting observables, without any empirical penalty forthcoming. Moreover, by interpreting joint probabilities in the SOBST framework, we know now that introducing joint probabilities for upward incompatible events amounts to constructing an instruction set for the possible outcomes of these events – a determinate-value hidden variable model subject to counterfactual definiteness.

Given the result of this section, the following question remains: In some cases, joint probabilities for non-commuting observables are empirically unproblematic. In other cases, most prominently in the case of the Bell/Aspect experiment, the definability of such joint probabilities in a model causes the model to be empirically inadequate. What is the differentiating feature – what is the common aspect of those setups for which joint probabilities lead to empirical trouble? We have seen that the fact that non-commuting observables are involved does not differentiate. Also, the number of observables involved is not a good measure: for the Bell/Aspect experiment with the direction setting fixed in one wing, joint probabilities cause no trouble.[31]

We believe that an easy answer to this question is not to be had. It should be clear that intricate details of the setup will be decisive. Thus, if joint probabilities cause empirical trouble for a certain setup, this signals something about the physics of the setup, not just about mathematics.

6. FACTORIZABLE STOCHASTIC HIDDEN VARIABLES

We have already shown that the minimalist interpretation cannot be sustained. Yet, here is a moral to be drawn from Fine's theorem: do not put all the probabilities of the Bell/Aspect experiment in a single large probability space. In the light of the preceding section, this means that determinate-value models, in which a global joint probability measure p_{E_1,E_2,F_1,F_2} is definable, are out. Suppose thus one sets out to analyze the experiment, consistently using only the four probability measures $\mu_{E_1 \cup F_1}$, $\mu_{E_1 \cup F_2}$, $\mu_{E_2 \cup F_1}$, and $\mu_{E_2 \cup F_2}$ that are defined on the families of outcomes of the corresponding events. In order to be explanatorily adequate, the analysis should respect Screening-off and the Locality Requirement (cf. Section 2.2). In hidden variable terminology, the task is therefore to construct a factorizable stochastic h.v. model.

Figure 4 represents a factorizable stochastic hidden variable model of the Bell/Aspect experiment. The model contains additional ('hidden') structure (a common common cause event C with K atomic outcomes)[32] that is not present in the surface description of the experiment (cf. Figure 2). The new model must reproduce the given surface probabilities (condition of Adequacy). The additional structure is introduced in order to make the resulting model explanatorily satisfying; it has to respect the con-

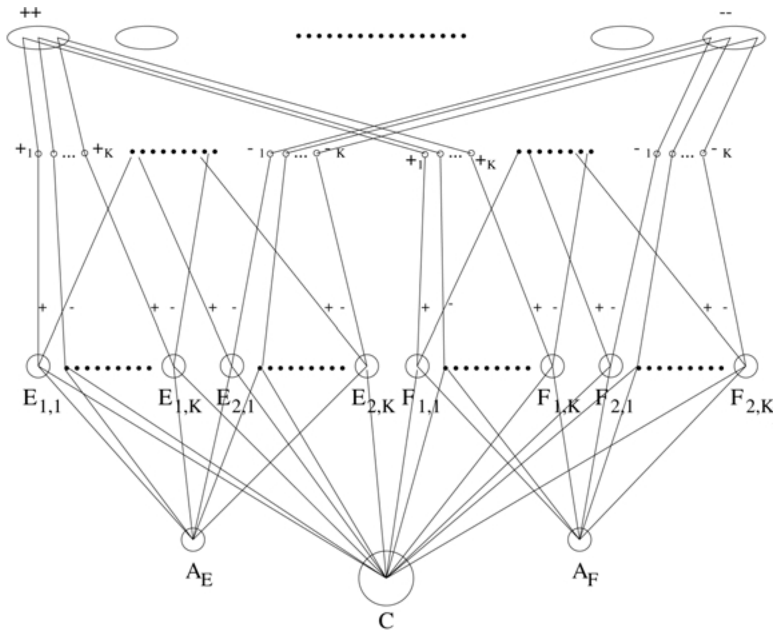

Figure 4. SOBST model of the Bell/Aspect experiment with factorizable stochastic hidden variables. The K-multiplied extended outcomes are only shown for the outcome '$++$' of $E_1 \cup F_1$ and for outcome '$--$' of $E_2 \cup F_2$ in order to keep the picture tractable. Ovals on the top correspond to surface outcomes, i.e., they integrate outcomes differentiated through the outcomes of the common common cause C.

straints of Screening-off and the Locality Requirement. We have already motivated these constraints in Section 2.2. With respect to the Bell/Aspect experiment, Screening-off embodies a conviction about how correlations can be (causally) explained. The ('hidden') Locality Requirement has two sides. First, it allows one to deduce the usual demand for Locality:

$$\text{(14)} \quad \begin{aligned} \overset{*}{\mu}_{E_i \cup F_1 \cup C}(e_i \cap \mathbf{1}_{F_1} \cap \omega_k) &= \overset{*}{\mu}_{E_i \cup F_2 \cup C}(e_i \cap \mathbf{1}_{F_2} \cap \omega_k) \\ \overset{*}{\mu}_{E_1 \cup F_j \cup C}(\mathbf{1}_{E_1} \cap f_j \cap \omega_k) &= \overset{*}{\mu}_{E_2 \cup F_j \cup C}(\mathbf{1}_{E_2} \cap f_j \cap \omega_k) \end{aligned}$$

This condition makes sure that what happens in one wing is independent of what happens in the other wing and thus guarantees that the distant correlations are really explained in a local way: if you allow for nonlocality in a 'hidden' model in order to explain puzzling nonlocal surface correlations, you can be accused of a combination of question-begging and mysticism – if you are willing to accept nonlocality, why not rest satisfied with the surface description? Secondly, the Locality Requirement yields the 'No Conspiracy' constraint, which reads

$$\text{(15)} \quad \overset{*}{\mu}_C\ (\omega_k) = \overset{*}{\mu}_{E_i \cup F_j \cup C}\ (\mathbf{1}_{E_i} \cap \mathbf{1}_{F_j} \cap \omega_k).$$

This condition rules out influences from the direction settings in the two wings on the particle source. It thus embodies the conviction that while our model should explain the surface probabilities, it *must not* explain the direction settings – otherwise, the model would be close to global determinism: an option that is always open logically, but that does not explain anything.

The constraints imposed on the model of Figure 4 are widely recognized as well motivated. However, in the model the Bell/CH inequalities are derivable.[33] Thus, while the model would be accepted as explanatorily sufficient, it is empirically inadequate: the surface probabilities are subject to the Bell/CH inequalities, which have been found to be empirically violated.

Fine's theorem suggests that the model of Figure 4 should allow for the definition of a joint probability measure p_{E_1,E_2,F_1,F_2}. We will now show how this joint measure can in fact be defined. The definition is exactly like Fine's, and should come as no surprise. Still, we will show explicitly how all the features used to specify the model play a crucial role in the definition.

We thus define the joint probability measure[34] as

$$\begin{aligned}(16)\quad & p_{E_1,E_2,F_1,F_2}(e_1, e_2, f_1, f_2)\\ &= \sum_{i=1}^{K} \overset{*}{\mu}_{E_1\cup C}\ (e_1 \cap \mathbf{1}_C \mid \omega_i)\cdot \overset{*}{\mu}_{E_2\cup C}\ (e_2 \cap \mathbf{1}_C \mid \omega_i)\cdot\\ &\qquad \overset{*}{\mu}_{F_1\cup C}\ (f_1 \cap \mathbf{1}_C \mid \omega_i)\cdot \overset{*}{\mu}_{F_2\cup C}\ (f_2 \cap \mathbf{1}_C \mid \omega_i)\cdot \overset{*}{\mu}_{C}\ (\omega_i).\end{aligned}$$

In this definition, we made use of the Locality Requirement in order to define, e.g.,

$$\begin{aligned}(17)\quad \overset{*}{\mu}_{E_1\cup C}\ (e_i \cap \omega_j) &= \overset{*}{\mu}_{E_1\cup F_1\cup C}\ (e_i \cap \mathbf{1}_{F_1} \cap \omega_j)\\ &= \overset{*}{\mu}_{E_1\cup F_2\cup C}\ (e_i \cap \mathbf{1}_{F_2} \cap \omega_j).\end{aligned}$$

Appendix B shows that the measure (16) indeed returns the given single and joint surface probabilities. Significantly, in this proof we need no more and no less than the entire set of all the conditions of the common common cause and all the conditions imposed on the SOBST model that also allowed for the derivation of the Bell/CH inequalities. If one of the conditions is dropped, the joint measure is not consistently definable.[35]

Note that the joint probability p_{E_1,E_2,F_1,F_2} of Equation (16) is *not* an empirical probability in the model – it is defined purely mathematically as a weighed average over the probabilities of hidden states, not as the probability of hidden states themselves. Nevertheless, by the construction of Section 5, we can construct a determinate-value (fatalistic) model, in which p_{E_1,E_2,F_1,F_2} is the probability of outcomes of an event C providing instruction sets. We will comment on this in Section 7 below.

The moral to be drawn from this section is simple: the prudent strategy of working with four small probability spaces fails. Given that the model is constrained by the physically motivated conditions of Adequacy, Screening-off, and the Locality Requirement, the initially disconnected small probability spaces can be pasted together to form one large probability space in which the troublesome measure p_{E_1,E_2,F_1,F_2} is defined.

7. DETERMINATE-VALUE VS. STOCHASTIC HIDDEN VARIABLES

A question that remains is the relation between determinate-value models discussed in Section 5 and factorizable models discussed in Section 6. Is there any generality achieved by moving from the former to the latter model?

We will now show that for the Bell/Aspect experiment, if it is possible to construct a model of the first kind, then it is also possible to construct the model of the second kind, and the other way round. Starting with the stochastic model from Section 6, we showed that the global joint probability measure p_{E_1,E_2,F_1,F_2} is definable. The event C in the stochastic model is required to be a common common cause and not a source of instruction sets. Thus, the number K of atomic outcomes of C does not have to coincide with the number of possible instruction sets. However, preserving the location of C in the model, we can turn the purely mathematical probability p_{E_1,E_2,F_1,F_2} into an empirical probability and give an alternative SOBST model of the situation in which the event C has as many atomic outcomes as there are quadruples of atomic outcomes of the E_i, F_j, i.e., 16 in our example. Thus, the mathematical definability of the joint measure p_{E_1,E_2,F_1,F_2} allows for the construction of a determinate-value h.v. model. – In the other direction, given a factorizable determinate-value h.v. model (i.e., a model with an event C with 16 atomic outcomes $\omega_{\pm,\pm,\pm,\pm}$ corresponding to the possible instruction sets), we can simply read off the joint probability measure:

$$(18) \qquad p_{E_1,E_2,F_1,F_2}(e_1, e_2, f_1, f_2) = \mu_C(\omega_{e_1,e_2,f_1,f_2}).$$

In this way, we have given a SOBST interpretation of one more of the equivalences Fine has established:[36]

> *The probability measure p_{E_1,E_2,F_1,F_2} can be defined to return the given joint probabilities $\mu_{E_\cup,F_j}$, $i, j = 1, 2$, of the setup as marginals if and only if we can give a factorizable determinate-value hidden variable model for the setup.*

The implication to the right means that if you can produce a factorizable stochastic model, then you can trivialize it, in the sense of giving a determinate-value model for the same surface probabilities. This may come as a surprise, given the philosophical rationales for preferring the stochastic models. In determinate-value (fatalistic) models, the probabilities $\mu_C(\omega_i)$ of outcomes of the common cause event C are generally nontrivial, but they can naturally be interpreted epistemically, thus allowing for an ignorance interpretation of the statistical character of QM. The conditional probabilities of outcomes of the events E_i and F_j, conditioned on atomic outcomes ω_k of C, are either 0 or 1. It is exactly the feeling that the determinism embodied in fatalistic models might be too restrictive to allow for reproducing the experimental data, that motivates the construction of stochastic models. By allowing for two levels of modality, i.e., two kinds of nontrivial probabilities (on the outcomes of the common cause and on the measurement results conditional on such outcomes), it is hoped that the class of surface data that can be modeled will be enlarged considerably. This hope is frustrated, as the equivalence shows. Also on a purely mathematical level, the equivalence is surprising, since for a model to be stochastic, it has to fulfill the Screening-off Equations (3), whereas a determinate-value model has to fulfill these equations with trivial conditional probabilities, e.g., $\mu_{E\cup F\cup C}(e \mid \omega_i) \in \{0, 1\}$ – a much more stringent requirement. The demand for Screening-off itself appears not too restrictive.

Contrary to these appearances, the above equivalence shows that given the constraints imposed on the stochastic model, the hidden variable construction can be reiterated (via the definition of the global joint probability measure p_{E_1,E_2,F_1,F_2}), collapsing the two levels of modality into one without imposing any new constraints on the surface probabilities that the model is able to reproduce. Contrary to what was intended, stochastic models do not embody two genuine levels of modality.

A closer look at the situation reveals that for a proponent of stochastic hidden variables, the situation is actually worse than that; most setups do not allow for truly stochastic models (i.e., stochastic models with nontrivial conditional probabilities). Even barring the physically motivated argument by Suppes and Zanotti to the effect that one should not consider stochastic models in the first place,[37] it is a mathematical fact that only very few situations allow for the satisfaction of the Screening-off equations with non-trivial conditional probabilities. You cannot even satisfy these equations in a two-wing setup with three atomic outcomes in each wing, if you do not allow for trivial conditional probabilities.[38]

Thus, contrary to whatever pushed people into constructing stochastic models, such models are either more restrictive than determinate-value models (if they are truly stochastic), or amount to the same (if they are stochastic in the broad sense) – the condition of Screening-off is in fact very strong.

8. CONCLUSIONS

Our aim in this paper was to show that the minimalist interpretation of the violation of Bell's inequalities is wrong: the reason why Bell's inequalities are violated is not to be identified with the non-commutativity of the quantum observables involved. We first argued that (i) no argument is known supporting the claim that observables that QM classifies as incompatible cannot have a joint probability distribution, and (ii) the claim is actually false. Thus, the argument in favor of the minimalist interpretation that uses Fine's theorem is a non-starter, since its premise is missing. Secondly, we explicitly constructed in the SOBST framework joint probabilities for outcomes of incompatible events, and observed that a model for two events can be made empirically adequate. We saw that the introduction of empirical joint probabilities for incompatible events in a SOBST model amounts to introducing a determinate-value hidden variable model subject to counterfactual definiteness. As is implied by Fine's theorem, the definability of joint probabilities for all the observables in the Bell/Aspect experiment causes a model to obey the Bell/CH inequalities, and thus, to be empirically inadequate. In a third step, therefore, we read the equivalence of Fine's theorem backward, arriving at a guide to strategy: an adequate model for the Bell/Aspect experiment must not define joint probabilities for all observables. This strategy, embodied by the class of factorizable stochastic models, was however seen to fail: the physically motivated constraints of Screening-off and the Locality Requirement imposed on such models allowed us to paste the four initially disconnected probability spaces in the model into one large space, defining the troublesome joint probabilities. In the fourth and final step, we discussed whether stochastic models are less restrictive than determinate-value models, coming to the conclusion that, surprisingly, truly stochastic factorizable models are much more restrictive than deterministic factorizable models.

The upshot of our paper is that (1) the minimalist interpretation is wrong in claiming that the violation of Bell's inequalities is a purely mathematical fact, and (2) it is really the physically motivated constraints of Screening-off, Locality and No Conspiracy imposed on hidden variable models that are to blame for Bell's inequalities. Thus, the majority view is right: the violation of Bell's inequalities teaches a lesson at least about explanation, if not about metaphysical issues. Bell's inequalities belong to physics, not just to mathematics.

APPENDIX A. COMMON CAUSES IN THE SOBST FRAMEWORK

The idea of a common cause as advocated by (Reichenbach, 1956) has already been motivated in the discussion of Screening-off in Section 2.2: correlations between outcomes of space-like separated events E and F are (causally) explained if and only if you can give an adequate extended SOBST model (i.e., one with the same surface probabilities) in which there is a common cause C in the common past of E and F such that conditional on the atomic outcomes of C, the correlations disappear. (Thus, the idea is roughly that the population in which the outcomes of E and F are correlated can be partitioned (through the atomic outcomes of C) in such a way that in each of the partitions, the correlations vanish.)

Technically, the common cause event C will have K atomic outcomes. The events E and F and their atomic outcomes e_i $(i \leq I)$ and f_j $(j \leq J)$ will be replaced by K copies E_k, F_k, e_i^k, and f_j^k, $k \leq K$, such that in each atomic outcome ω_k of C there will be just the corresponding (space-like separated) events E_k and F_k and the corresponding atomic outcomes e_i^k and f_j^k of E_k and F_k. Furthermore, it is required that for $i \leq I$, $j \leq J$ and $k, l \leq K$,

$$(19) \qquad e_i^k \cap f_j^l \neq \emptyset \qquad \text{iff} \quad k = l.$$

In the model thus extended, there is now new freedom to adjust the (hidden) probabilities $\mu^*_{E_k \cup F_k \cup C}(e_i^k \cap f_j^k \cap \omega_k)$ in such a way that conditional on the ω_k, the correlations disappear. As the surface probabilities of the original model have to be reproduced, the condition of Adequacy demands that the hidden probabilities, summed over all the atomic outcomes of C, return the original surface probabilities.

In the situation of the Bell/Aspect experiment, we need to generalize the above idea of a common cause to what has been called a *common*

common cause.[39] Recall that the experiment is designed such that at the time the particle pair leaves the source, it is not yet established which of four possible correlations will be measured.[40] The common cause C located in the common past of the measurements (read: at the position of the particle source) should thus be able to account not just for one, but for all four types of correlations that can be observed (cf. Figure 2). Therefore, we define a common common cause as follows:[41]

DEFINITION 11 (of common common cause). Let $\{E_1, F_1\}$, $\{E_2, F_2\} \ldots \{E_N, F_N\}$ be a collection of space-like sets in a SOBST model $\langle W, \leq, \mathcal{E}, \Upsilon\rangle$. Let $E_n \cup F_n$ and $E_m \cup F_m$ be upward incompatible for all $n \neq m$ ($n, m \leq N$), and let events A_E and A_F exist such that $A_E \prec E_n$ and $A_F \prec F_n$ for all $n \leq N$. Further, let $e_{n;1}, \ldots, e_{n;I}$ be atomic outcomes of E_n and $f_{n;1}, \ldots f_{n;J}$ be atomic outcomes of F_n, such that $e_{n;i} \cap f_{n;j} \neq \emptyset$ for any $i \leq I$, $j \leq J$ and $n \leq N$. Let also, for each $n \leq N$ some outcomes of $E_n \cup F_n$ be correlated, i.e., for some (possibly all) $i \leq I$ and $j \leq J$:

$$\mu_{E_n \cup F_n}(e_{n;i} \cap f_{n;j}) \neq \mu_{E_n \cup F_n}(e_{n;i}) \cdot \mu_{E_n \cup F_n}(f_{n;j}).$$

A *common common cause* C that accounts for all these correlations is an event with K atomic outcomes ω_k in a model $\langle W^*, \leq^*, \mathcal{E}^*, \Upsilon^*\rangle$ that is enlarged in respect to $\langle W, \leq, \mathcal{E}, \Upsilon\rangle$ by:

1. K-extending all the $E_n - F_n$ structures;
2. adding an event C with K atomic outcomes ω_k;
3. requiring that for $n \leq N$, $i \leq I$, and $j \leq J$, all $e^k_{n;i}$ and all $f^k_{n;j}$ split in C, i.e.,
$$\forall x \in e^k_{n;i}, y \in e^l_{n;i}, k \neq l \; C(x, y) \cap C \neq \emptyset \quad \text{and}$$
$$\forall x \in f^k_{n;j}, y \in f^l_{n;j}, k \neq l \; C(x, y) \cap C \neq \emptyset;$$
4. defining measures $\overset{*}{\mu} \in \Upsilon^*$ that satisfy the conditions of Screening-off and Adequacy (below).

The common common cause C should satisfy:

For all $i \leq I$, $j \leq J$, $k \leq K$, $n \leq N$,

$$\overset{*}{\mu}_{E^k_n \cup F^k_n \cup C}(e^k_{n;i} \cap f^k_{n;j} \mid \omega_k)$$
$$= \overset{*}{\mu}_{E^k_n \cup F^k_n \cup C}(e^k_{n;i} \mid \omega_k) \cdot \overset{*}{\mu}_{E^k_n \cup F^k_n \cup C}(f^k_{n;j} \mid \omega_k), \text{ (Screening-off)}$$

where $\overset{*}{\mu}_{E^k_n \cup F^k_n \cup C}(\omega_k) = \overset{*}{\mu}_{E^k_n \cup F^k_n \cup C}(\mathbf{1}_{E^k_n} \cap \mathbf{1}_{F^k_n} \cap \omega_k)$;

For all $i \leq I$, $j \leq J$, $n \leq N$,

$$\mu_{E_n \cup F_n}(e_{n;i} \cap f_{n;j}) = \sum_{k=1}^{K} \overset{*}{\mu}_{E_n^k \cup F_n^k}(e_{n;i}^k \cap f_{n;j}^k). \qquad \text{(Adequacy)}$$

Events $E_n \cup F_n$ and $E_m \cup F_m$, $n \neq m$ are assumed to be upward incompatible, since they represent alternative measurements. As they are alternative, there is some even at which one or another measurement is chosen, and this translates into requiring that the events A_E and A_F exist.

9. PROOFS FROM SECTION 6

We will prove that the global joint probability measure p_{E_1,E_2,F_1,F_2} defined in equation (16) indeed returns the given single and joint surface probabilities as marginals. For the single probabilities μ_{E_i} and μ_{F_j}, note that, e.g. (the other three cases are exactly analogous),

$$(20) \qquad \overset{*}{\mu}_{E_1 \cup C}(\mathbf{1}_{E_1} \cap \mathbf{1}_C \mid \omega_i) = \frac{\overset{*}{\mu}_{E_1 \cup C}(\mathbf{1}_{E_1} \cap \mathbf{1}_C \cap \omega_i)}{\overset{*}{\mu}_{E_1 \cup C}(\mathbf{1}_{E_1} \cap \omega_i)} = 1$$

and, e.g.,

$$(21) \qquad \sum_{i=1}^{K} \overset{*}{\mu}_{E_2 \cup C}(e_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_C(\omega_i)$$

$$= \sum_{i=1}^{K} \overset{*}{\mu}_{E_2 \cup C}(e_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{E_2 \cup C}(\mathbf{1}_{E_2} \cap \omega_i)$$

$$= \overset{*}{\mu}_{E_2 \cup C}(e_2 \cap \mathbf{1}_C) = \overset{*}{\mu}_{E_2}(e_2) = \mu_{E_2}(e_2),$$

where the first equation is based on the Locality Requirement in the form of No Conspiracy, the second follows from the theorem of total probability, the third is a consequence of the Locality Requirement, and the last follows from the condition of Adequacy.

Thus, the posited measure p_{E_1,E_2,F_1,F_2} indeed returns the single probabilities as marginals. To see that it also returns the given joint probabilities for the four pairs of upward compatible events, let us calculate as an example (the other three cases are again analogous):

$$(22)\quad p_{E_1,E_2,F_1,F_2}(e_1, \mathbf{1}_{E_2}, \mathbf{1}_{F_1}, f_2)$$

$$= \sum_{i=1}^{K} \overset{*}{\mu}_{E_1\cup C}(e_1 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{F_2\cup C}(f_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_C(\omega_i).$$

By the Locality Requirement we can identify

$$(23)\quad \overset{*}{\mu}_{E_1\cup C}(e_1 \cap \mathbf{1}_C \mid \omega_i) = \overset{*}{\mu}_{E_1\cup F_2\cup C}(e_1 \cap \mathbf{1}_{F_2} \cap \mathbf{1}_C \mid \omega_i),$$

$$\overset{*}{\mu}_{F_2\cup C}(f_2 \cap \mathbf{1}_C \mid \omega_i) = \overset{*}{\mu}_{E_1\cup F_2\cup C}(\mathbf{1}_{E_1} \cap f_2 \cap \mathbf{1}_C \mid \omega_i).$$

Screening-off then gives us

$$(24)\quad \overset{*}{\mu}_{E_1\cup C}(e_1 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{F_2\cup C}(f_2 \cap \mathbf{1}_C \mid \omega_i)$$

$$= \overset{*}{\mu}_{E_1\cup F_2\cup C}(e_1 \cap f_2 \cap \mathbf{1}_C \mid \omega_i).$$

Next, the Locality Requirement in the form of 'No Conspiracy' guarantees:

$$(25)\quad \overset{*}{\mu}_C(\omega_i) = \overset{*}{\mu}_{E_1\cup F_2\cup C}(\omega_i) = \overset{*}{\mu}_{E_1\cup F_2\cup C}(\mathbf{1}_{E_1} \cap \mathbf{1}_{F_2} \cap \omega_i).$$

Accordingly, by the theorem of the total probability:

$$(26)\quad \sum_{i=1}^{K} \overset{*}{\mu}_{E_1\cup F_2\cup C}(e_1 \cap f_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{E_1\cup F_2\cup C}(\mathbf{1}_{E_1} \cap \mathbf{1}_{F_2} \cap \omega_i)$$

$$= \overset{*}{\mu}_{E_1\cup F_2\cup C}(e_1 \cap f_2 \cap \mathbf{1}_C),$$

and the Locality Requirement guarantees

$$(27)\quad \overset{*}{\mu}_{E_1\cup F_2\cup C}(e_1 \cap f_2 \cap \mathbf{1}_C) = \overset{*}{\mu}_{E_1\cup F_2}(e_1 \cap f_2),$$

from which, via Adequacy, we get in fact the desired result

$$(28)\quad p_{E_1,E_2,F_1,F_2}(e_1, \mathbf{1}_{E_2}, \mathbf{1}_{F_1}, f_2) = \mu_{E_1\cup F_2}(e_1 \cap f_2).$$

Thus, the posited joint measure indeed yields the 'double' joints. In the derivation above, observe how this fact is secured by the conditions of Screening-off, Adequacy, and the Locality Requirement, each of which had to be invoked for the derivation to go through.

In addition to returning the joint surface probabilities for pairs of upward compatible events, the measure p_{E_1,E_2,F_1,F_2} also allows for a coherent definition of joint probability measures for upward incompatible events, for which there are no surface μ's to compare them with. For instance,

$$(29)\qquad p_{E_1,E_2}(e_1,e_2) = p_{E_1,E_2,F_1,F_2}(e_1,e_2,\mathbf{1}_{F_1},\mathbf{1}_{F_2})$$

$$= \sum_{i=1}^{K} \overset{*}{\mu}_{E_1\cup C}(e_1 \cap \mathbf{1}_C \mid \omega_i)\cdot \overset{*}{\mu}_{E_2\cup C}(f_2 \cap \mathbf{1}_C \mid \omega_i)\cdot \overset{*}{\mu}_C(\omega_i).$$

ACKNOWLEDGEMENTS

Tomasz Placek is supported by KBN grant 191/H01/97/12. We would like to thank our audiences at the Oxford Philosophy of Physics seminar, April 23, 1999, at the LMPS conference, Cracow, August 22, 1999, and at Konstanz, November 10, 1999, for valuable discussions.

NOTES

[1] Cf., e.g., (Aspect et al., 1982; Weihs et al., 1998). For a critical assessment of this claim, cf., e.g., (Huelga et al., 1995). The experiment is pictured in Figure 2 below. In what follows, we will assume that (i) is correct, i.e., that the inequalities have in fact been violated.

[2] (Fine, 1982a; Fine, 1982b).

[3] (Pitowsky, 1989).

[4] We do not wish to insinuate that Fine or Pitowsky explicitly argued in favor of a minimalist interpretation, only that their results have been used to back such an interpretation. While Pitowsky's view seems quite close to a minimalist position (cf. (Pitowsky, 1989, pp. 8, 50, 180) and also note 9 below), Fine seems much less committed to a minimalist interpretation.

[5] (Fine, 1982a, p. 294).

[6] Pitowsky, who insists that all the experimental data must be embeddable in a single probability space of some sort, takes this result to teach us that we need to change our notion of probability: certain non-classical probability spaces do allow for an embedding. Cf., e.g., (Pitowsky, 1989, p. 50).

[7] (Kowalski and Placek, 1999; Placek, 2000).

[8] In particular, the finite probability spaces used in Bell-type arguments can always be thus represented. Cf. Theorem 2 below.

[9] (Belnap, 1992).

[10] (Kowalski and Placek, 1999).

[11] (Placek, 2000).

[12] This is the main algebraic result proved in (Kowalski and Placek, 1999).

[13] For a proof, cf. (Kowalski and Placek, 1999).

[14] Cf. (Kowalski and Placek, 1999, lemma 6).

[15] In line with the emphasis on algebraic structures inherent in our approach, we will specify probability spaces as pairs $\langle \mathfrak{B}, \mu \rangle$. Standard usage treats probability spaces as triples $\langle \Omega, \mathfrak{B}, \mu \rangle$, where Ω is the sample space and $\mathfrak{B}$ is a Boolean σ-algebra of subsets of Ω. In fact, the sample space is already uniquely specified through the unit of the Boolean algebra $\mathfrak{B}$, so that we do not need to mention the sample space explicitly.

[16] For a proof, cf. the Appendix of (Placek, 2000).

[17] In analyzing Bell's inequalities, we will need the concept of a common common cause, not just common causes as motivated here. For the somewhat lengthy definition of 'common common cause' in the SOBST framework and for the notion of an 'extended model', cf. Appendix A.

[18] A natural extension to countably many spaces is also possible, but we will not need to consider it here.

[19] Given an event E with associated probability space $\langle \mathfrak{B}_E, \mu_E \rangle$, we define a random variable f_E in the standard way as a measurable real-valued function on the atoms of $\mathfrak{B}_E$.

[20] The theorem is originally stated in terms of joint distributions for quantum observables (Fine, 1982a; Fine, 1982b). We give here an equivalent reading that only uses probability measures.

[21] For the definition of correlation polytopes, cf. (Pitowsky, 1989, chap. 2.3).

[22] He attributes his theorems (2.4) and (2.5) to (Fine, 1982a); cf. (Pitowsky, 1989, p. 50).

[23] Cf. (Szabó, 2000).

[24] We chose Fine's formulation because it is more frequently discussed in the current literature. For more general considerations, Pitowsky's formulation would be a better starting point, since the correlation polytope approach reveals the geometrical background of the inequalities in an easily generalizable way.

[25] Cf., e.g., (De Beare et al., 1999, p. 69): "the nonvanishing of a commutator between two observables is equivalent to the nonexistence of a jpd [joint probability distribution] for these observables."

[26] Cf. again (De Beare et al., 1999, p. 69): "(non)locality is irrelevant for accounting for the conflict between BI [Bell's inequalities] and QM."

[27] For a review of the relevant theorems, see (Bugajski, 1976; Bugajski, 1978).

[28] Cf., e.g., Fine's Theorem 7 (Fine, 1982b, p. 1309) and Nelson's Theorem 14.1 (Nelson, 1967, p. 117).

[29] The theorems, on the other hand, are of course correct, but they establish something else, namely that defining joint probabilities for non-commuting observables amounts to leaving the formalism of QM. In our view, this tells against using Fine's (jd) condition (Fine, 1982b) that links the existence of joint distributions with commutativity of the involved operators, as (jd) is too strong: many joint probabilities not allowed by (jd) may peacefully be defined as purely mathematical probabilities. For a similar point, cf. (Svetlichny et al., 1988).

[30] Cf. the discussion of section 4 above, esp. Section 4.2.

[31] Cf. (Landau, 1987; Svetlichny et al., 1988).

[32] Cf. Appendix A for details on the notion of a common common cause employed here as well as for the notion of an extended model. Below, μ^* denotes probability measures defined in the extended model.

[33] For a proof in the SOBST framework, cf. (Placek, 2000).
[34] Cf. (Fine, 1982a, eq. (3)).
[35] In the surface model of Figure 2, the joint measure is not definable – as is to be expected from Fine's theorem, since the surface probabilities violate the Bell/CH inequalities.
[36] (Fine, 1982a). He uses 'deterministic' for 'factorizable determinate-value'.
[37] (Suppes and Zanotti, 1976).
[38] For a proof, cf. (Placek, 2000).
[39] Cf. (Hofer-Szabó et al., 1999).
[40] The recent experiment reported by (Weihs et al., 1998) gives strong support to this reading.
[41] It is arguable whether this constitutes a genuine extension of Reichenbach's original idea of a common cause. At any rate, Reichenbach has not formulated his common cause principle in a way that would be applicable to the Bell/Aspect experiment, so that we need to be at least more explicit about the condition.

REFERENCES

Aspect, A., Dalibard, J., and Roger, G.: 1982, 'Experimental Test of Bell's Inequalities Using Time-Varying Analyzers', *Physical Review Letters* **49**, 1804–1807.

Belnap, N.: 1992, 'Branching Space-Time', *Synthese* **92**, 385–434.

Bugajski, S.: 1976, 'Joint Probabilities in Quantum Mechanics', Unpublished typescript, Katowice.

Bugajski, S.: 1978, 'A Remark on Joint Probability Distributions for Quantum Observables', *Zeitschrift für Naturforschung* **33a**, 1383–1385.

De Beare, W., Mann, A., and Revzen: M.: 1999, 'Locality and Bell's Theorem', *Foundations of Physics* **29**, 67–77.

Fine, A.: 1982a, 'Hidden Variables, Joint Probability, and the Bell Inequalities', *Physical Review Letters* **48**, 291–295.

Fine, A.: 1982b, 'Joint Distributions, Quantum Correlations, and Commuting Observables', *Journal of Mathematical Physics* **23**, 1306–1310.

Hofer-Szabó, G., Rédei, M., and Szabó, L. E.: 1999, 'On Reichenbach's Common Cause Principle and Reichenbach's Notion of Common Cause', *British Journal for the Philosophy of Science* **50**, 377–399.

Huelga, S., Ferrero, M., and Santos, E.: 1995, 'Loophole-free Test of the Bell Inequality', *Physical Review A* **51**, 5008–5011.

Kowalski, T. and Placek, T.: 1999, 'Outcomes in Branching Space-Time', *British Journal for the Philosophy of Science* **50**, 349–375.

Landau, L. J.: 1987, 'On the Violation of Bell's Inequality in Quantum Theory', *Physics Letters* **A120**, 54–56.

Nelson, E.: 1967, *Dynamical Theories of Brownian Motion*, Princeton University Press, Princeton, NJ.

Pitowsky, I.: 1989, *Quantum Probability – Quantum Logic*, Springer, Berlin.

Placek, T.: 2000, 'Stochastic Outcomes in Branching Space-Time: Analysis of Bell's Theorem', *British Journal for the Philosophy of Science* **51**(3), 445–475.

Reichenbach, H.: 1956, *The Direction of Time*, University of California Press, Berkeley.

Suppes, P. and Zanotti, M.: 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*, D. Reidel, Dordrecht, pp. 445–455.

Svetlichny, G., Redhead, M., Brown, H., and Butterfield, J.: 1988, 'Do the Bell Inequalities Require the Existence of Joint Probability Distributions?', *Philosophy of Science* **55**, 387–401.

Szabó, L. E.: forthcoming 2000, 'Critical Reflections on Quantum Probability Theory', in M. Rédei and M. Stoelzner (eds), *John von Neumann and the Foundations of Physics*, Kluwer, Dordrecht.

Weihs, G., Jennewein, T., Simon, C., Weinfurter, H., and Zeilinger, A.: 1998, 'Violation of Bell's Inequality under strict Einstein Locality Conditions', *Physical Review Letters* **81**, 5039–5043.

Thomas Müller
Philosophisches Seminar I
Universität Freiburg
D–79085 Freiburg
Germany
E-mail: tmueller@fdm.uni-freiburg.de

Tomasz Placek
Department of Philosophy
Jagiellonian University
Grodzka 52
PL 31-044 Cracow
Poland
E-mail: uzplacek@cyf-kr.edu.pl