On the Issue of Joint Probabilities in Interpretations of Bell's Inequalities*

Thomas Müller, Tomasz Placek

September 23, 1999

Contents

*We would like to thank our audiences at the Oxford Philosophy of Physics seminar, April 23, 1999, and at the LMPS conference, Cracow, August 22, 1999, for valuable discussions. Authors' addresses: Thomas Müller, Philosophisches Seminar I, Universität Freiburg, D–79085 Freiburg, Germany; *e-mail*: `tmueller@fdm.uni-freiburg.de`; Dr. Tomasz Placek, Department of Philosophy, Jagiellonian University, Grodzka 52, PL 31-044 Cracow, Poland, *e-mail*: `uzplacek@cyf-kr.edu.pl`. T.P. is supported by KBN grant 191/H01/97/12.

1 Introduction

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 standard 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]

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 as Fine himself stresses, 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.

[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 3 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 be used to back such an interpretation. While Pitowsky's view seems quite close to a minimalist position (cf. also note 6 below), Fine seems much less committed to a minimalist interpretation. — For a strong statement of the minimalist view, cf. (De Muynck, 1986).

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

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 clarify the role of joint probabilities in the Bell/Aspect experiment, which is of crucial importance in the arguments in favor of the minimalist interpretation. We will 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. In order to proceed with formal rigor, we will use the framework of 'stochastic outcomes in branching space-time' (SOBST),[7] which allows for a unified treatment of causal, spatiotemporal and modal relations. SOBST models appear particularly suited for a discussion of joint probabilities in quantum mechanics, since most classical probability spaces have a representation in models of SOBST.[8] Furthermore, while we do not question the mathematics of Fine's and Pitowsky's theorems, we will draw a distinction, similar to that used by (Svetlichny et al., 1988), between purely mathematical joint probabilities and empirical joint probabilities. This distinction is a natural by-product of SOBST: roughly, joint probabilities are empirical iff there is in a model of SOBST a *properly located* event, whose outcomes constitute a probability space isomorphic to the probability space on which a given joint probability is defined.

In a second step, we will construct a hidden variable model for the Bell/Aspect experiment in the framework of generalized probability theory. This model appears to vindicate the minimalist interpretation as it (1) does not assign joint probabilities to results of non-commuting observables and (2) it is not committed to Bell's inequalities. However, as we will show, models based on generalized probability theory inevitably come into conflict with the predictions of quantum mechanics regarding probabilities of results of certain experiments.

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 contrasts the mathematical results of section 3 with the interpretation of joint probability spaces in the SOBST framework. Section 6 is the application of the SOBST analysis of joint probabilities to two models of the Bell/Aspect experiment. Section 7 treats interpretational issues and shows that the minimalist interpretation faces insuperable difficulties. Finally, section 8 discusses hidden variable models based on generalized probability theory. Two appendices contain technical material from section 6.

[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), and see section 8 below.

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

[8] In particular, the finite probability spaces used in Bell-type arguments can always be thus represented. Cf. (Kowalski and Placek, 1999, lemma 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 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 **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* $\mathcal{H}$.

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 non-empty 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 $C(x, y) \supseteq C(z, t)$.

[9] (Belnap, 1992).
[10] (Kowalski and Placek, 1999).
[11] (Placek, 1999).

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 8 (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* $i, j \leq n$

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

[12] This is the main algebraic result proved in (Kowalski and Placek, 1999).
[13] For a proof, cf. (Kowalski and Placek, 1999).

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

$$\mu_E(\mathbf{1}_E) = 1 \text{ and for mutually disjoint } e_i \in \mathfrak{B}_E : \mu_E(\bigcup_{i=1}^{\infty} e_i) = \sum_{i=1}^{\infty} \mu_E(e_i).$$

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]

LEMMA 2
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, $\mathcal{E}$ — 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

[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 $\langle \mathfrak{B}_E, \mu_E \rangle$. 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 Boolean algebra structure $\mathfrak{B}$, so that we do not need to mention the sample space explicitly.

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 3

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. Solid 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 should serve to illustrate lemma 3.

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

The Locality Requirement If two events E and F with finitely many outcomes each are such that for any two non-empty outcomes e and f of E and F, resp., the intersection $e \cap f$ is non-empty, then the probabilities of the joint event $E \cup F$ must return the single probabilities as marginals:

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

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 Requirement has two sides. For surface models like the model of figure 3 below, in which all the probabilities are empirically established, the 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 4 below, the status of the 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 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 6.2 below).

[16] For a proof, cf. the Appendix of (Placek, 1999).

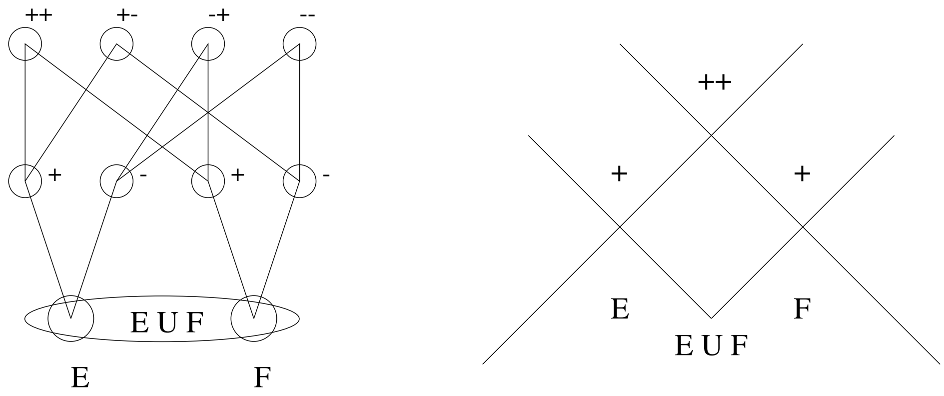

Figure 1: SOBST model with two upward compatible events E and F. Left: atomic outcomes '$+$' and '$-$' of E and F as well as the atomic outcomes '$++$', '$+-$', '$-+$', and '$--$' of the event $E \cup F$ are shown as a SOBST graph. Right: the relation between the atomic outcomes of the events E and F 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.

Screening-off The idea of Screening-off was introduced by (Reichenbach, 1956) in order to provide an interpretation of causality in an indeterministic framework. The SOBST reading of Reichenbach's idea is that correlations between space-like separated events are explained only if in joining the events, the probabilities factor, i.e., for space-like separated E and F,

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

It is exactly the occurrence of correlations between space-like separated events, i.e., violations of (2), that sends people looking for hidden variable (h.v.) models, i.e., for a common cause for the correlation: in an extended model with a common cause event C, (2) holds for outcomes differentiated by the outcomes ω_i of the common cause, i.e.,

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

2.3 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 (cf. figure 3 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.

[17] For the somewhat lengthy definition of 'common cause' in the SOBST framework and for the notion of an 'extended model', cf. Appendix A. — Section 7.2 below will show that Screening-off is a much stronger criterion than one might expect.

The restriction to finitely many outcomes allows us to keep to a simple notation in line with (Placek, 1999). 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 link to random variables, and so far we have not introduced random variables in the SOBST framework. While the introduction of random variables f_E in the SOBST framework poses absolutely no problem,[18] 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 '$-$' 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 Joint probability structures

In this section, we will review some mathematical probability theory. After comments on joint probabilities in QM in section 4, section 5 will indicate the connection between the mathematics and SOBST models. 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 mathematically defined probability measures, saving the letter μ for probability measures that have a SOBST interpretation in terms of weights on outcomes of events.

3.1 The mathematics of joint probabilities

Mathematically, 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. We believe that SOBST models allow for a thorough clarification of this issue. 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 $\mathcal{A}\times\mathcal{B} = \langle \mathfrak{B}_{A,B}, p_{A,B}\rangle$ with arbitrary probability measures $p_{A,B}$. As a restriction on the probability measure $p_{A,B}$, it is natural to demand that $p_{A,B}$ return the single measures p_A and p_B as marginals, i.e.,

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

This restriction, called the *marginal property*, is easy to fulfill in classical probability theory — in fact, there are generally infinitely many such measures. For 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 $p_{A,B}(\langle a,a\rangle) = p_{A,B}(\langle b,b\rangle) = r$, $p_{A,B}(\langle a,b\rangle) =$

[18] 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 $\mathfrak{B}_E$.

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

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

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

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

whereas with other measures, such correlations will occur. — As the mathematical method outlined here builds one probability space from two, it can be iteratively extended to any finite number of probability spaces.[19]

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 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 above. In a watered-down version (which is sufficient for our purposes here), Fine's theorem states the following:[20]

THEOREM 2

Given four 'single' probability spaces $\mathcal{A}_i, \mathcal{B}_j, i, j = 1, 2$, each with two atoms '$+$' and '$-$', and four 'double' probability spaces $\mathcal{A}_i \times \mathcal{B}_j$, $i, j = 1, 2$, such that the 'double' measures return the 'singles' 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 'double' (and thus* a fortiori *the four 'single') probabilities as marginals.*

2. *The probability measures on the 'single' and 'double' spaces satisfy the following four (Bell/CH) inequalities*

$$\begin{aligned} -1 \quad \leq \quad & p_{A_i,B_j}(+,+) + p_{A_i,B_{j'}}(+,+) + p_{A_{i'},B_{j'}}(+,+) \\ & - p_{A_{i'},B_j}(+,+) - p_{A_i}(+) - p_{B_{j'}}(+) \quad \leq \quad 0 \end{aligned} \tag{7}$$

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

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

[20] The theorem is originally stated in terms of joint distributions for quantum observables. We give here an equivalent reading that is somewhat simpler. Cf. section 2.3 above for the issue of statistical observables and random variables in our framework.

3.3 Pitowsky's theorem

Already formally, Pitowsky's theorem is similar to Fine's theorem: it links the fulfillment of certain inequalities to the possibility of treating given numbers (read: probabilities) as defined on a single joint probability space.

THEOREM 3
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* $S = \{\{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}\}$.[21]
2. *There exists a (classical) probability space* $\langle \mathfrak{B}, p \rangle$ *and* $A_1, A_2, A_3, A_4 \in \mathfrak{B}$ *such that*

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

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

$$\begin{aligned} &-1 \leq p_{ij} + p_{ij'} + p_{i'j'} - p_{i'j} - p_i - p_{j'} \leq 0, \\ &p_{ij} \geq 0, \quad p_i \geq p_{ij}, \quad p_j \geq p_{ij}, \quad p_i + p_j - p_{ij} \leq 1 \end{aligned} \tag{9}$$

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

3.4 Fine's and Pitowsky's theorems are equivalent

The parallel between equations (7) and (9) is striking — just read $p_{A_i} = p_i$, $p_{A_i,B_j} = p_{ij}$ etc. Note that (7) could easily be supplemented by the conditions apparently lacking for full equivalence: the definition of p_{A_i,B_j} as joint probabilities and p_{A_i} and p_{B_j} as marginals in fact guarantees that the second line of (9) is also fulfilled. Thus, the two theorems are equivalent. Pitowsky is explicit about this equivalence, attributing his theorems (2.4) and (2.5) to (Fine, 1982a).[22] Still, since the two theorems are often discussed separately, it might have been worth while to point out the equivalence again.

In what follows, we will always discuss Fine's theorem; the implications for Pitowsky's result should be obvious.[23]

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 are not violated, then a joint probability distribution for all the observables in the Bell/Aspect experiment is definable. It is, however, a well-entrenched view that QM prohibits the existence of joint probability distributions for non-commuting observables. Thus, the minimalist interpretation, via *modus tollens*, argues that the Bell/CH inequalities must be violated quite independently of any considerations about hidden variable models.

[21] For the definition of correlation polytopes, cf. (Pitowsky, 1989, chap. 2.3).
[22] Cf. (Pitowsky, 1989, p. 50).
[23] 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.

As a first step toward dismounting the minimalist interpretation, we will now 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.

It seems to be a general conviction that 'quantum mechanics teaches that there can be no joint probability distributions for non-commuting observables.' In this section, we will assess this claim. We will consider three arguments of increasing strength.

4.1 Non-commuting observables are not comeasurable

It is commonly held that in QM, observables corresponding to non-commuting operators cannot be measured simultaneously. Can you already argue against joint probabilities for non-commuting observables from the premise that non-commuting observables are not comeasurable? Let us assume that the two measurements are characterized by probability spaces $\langle \mathfrak{B}_1, p_1 \rangle$ and $\langle \mathfrak{B}_2, p_2 \rangle$, respectively, and 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. The problem, remember, is 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 single probabilities, e.g., $p(h_1) = 0$, as the probability that *in this kind of experiment, i.e., with two coins simultaneously tossed*, 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 obtaining a result on one 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 simultaneously two measurements 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 observables cites a simple matter of fact: The probabilistic algorithm of QM does not define such distributions. The algorithm allows for introducing joint probabilities for measurement results of pairs of commuting observables, but it does not work 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 the probability is:

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

This expression makes sense only for commuting A and B, because only commuting observables resolve into projectors such that any $P_A(e)\, P_B(f)$ 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 closed subspaces of an appropriate Hilbert space $\mathcal{H}$. Thus, the family of experimental propositions about a QM system is characterized by the set of closed subspaces of $\mathcal{H}$, 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" — we will consider such structures in some detail in section 8 below. These probability spaces should be used to define joint probabilities for QM observables. Now, classical joint probabilities satisfy certain natural properties, e.g., the marginal property (cf. eq. (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.[24] The immediate moral of this fact is that in standard QM, joint probabilities for non-commuting observables *cannot* be defined (in addition to the simple "are not defined" of the last argument) . Moreover, 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 distributions for non-commuting observables", this will suffer the same drawback

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

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 Assessment of the arguments

We have produced three arguments against joint probabilities for non-commuting QM observables, ordered by increasing strength: the first argument was 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, and this commitment is the largest in the case of the last and strongest argument.

To see how this commitment works, and to ultimately assess the significance of Fine's theorem, suppose that you believe in the empirical adequacy of QM, but that you remain agnostic as to the truth of its more abstract layer. More precisely, you believe that QM makes correct predictions about the possible outcomes of measurements and about the probabilities of these outcomes; however, you shrug your shoulders at a suggestion of any truth being in, say, the Hilbert space formalism.[25] If this is your position, you take the second argument as simply stating a fact about the probabilistic algorithm of QM, and the last argument does not cut any ice with you at all. Thus, if you are confronted with the general question whether there are joint probabilities for outcomes of measurements of noncommuting observables of a single quantum system (e.g., measurements of two non-orthogonal spin projections), you will probably take no stance. But now, given your empiricist convictions, Fine's theorem and the empirical violation of the Bell/CH inequalities are a revelation. They show that independently of the future development of physics, the two observables cannot have a joint probability distribution. Thus, despite of your initial reluctance, you are somewhat committed to the abstract layer of QM after all.

5 Joint probabilities in SOBST models

As a next move against the minimalist interpretation, and also as a s strategy that gives a new insight into joint probabilities in QM, we are going to interpret the joint probabilities of Fine's theorem as probabilities in some space-time structures. To this end, we first need to clarify the issue of joint probabilities in SOBST models.

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 $p_{E,F}$ 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 $\mathcal{E} \times \mathcal{F}$?

5.1 The joint space $\mathcal{E} \times \mathcal{F}$

From lemma 2 we know that whenever there is a (classical) probability space $\mathcal{A} = \langle \mathfrak{B}_A, p_A \rangle$ with $\mathfrak{B}$ complete and atomic (conditions which are always satisfied in the fi-

[25]It is not hard to take this position, given the interpretational problems of QM.

nite 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, lemma 2 is of little help. We will now show how to construct a unified model for the events E, F, and C. 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:

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

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. Moreover, it is this construction that we need in familiar applications of joint probabilities to description of multiple chance experiments, like multiple coin tosses or castings of dice. As an example, the joint probability $p(r_1, r_2)$ of obtaining result r_1 on the first coin and result r_2 on the second coin can be interpreted as the probability $\mu_{E_1 \cup E_2}(e_1 \cap e_2)$ of the event $E_1 \cup E_2$ of tossing the two coins, where e_i is the atomic outcome of E_i 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, and that we are given a probability measure $p_{E,F}$ that returns the measures μ_E and μ_F as marginals.

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 also 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, $C \cup E$ and $C \cup F$ 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.:

$$\begin{aligned} \mu_{E \cup C}(e \cap \mathbf{1}_C) = \mu_E(e), \quad \mu_{E \cup C}(\mathbf{1}_E \cap \omega) = \mu_C(\omega), \\ \mu_{F \cup C}(f \cap \mathbf{1}_C) = \mu_F(f), \quad \mu_{F \cup C}(\mathbf{1}_F \cap \omega) = \mu_C(\omega). \end{aligned} \tag{12}$$

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 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, 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

$$\begin{aligned} \mu_E(x) &= \mu_C(\omega_{x,+}) + \mu_C(\omega_{x,-}), \\ \mu_F(y) &= \mu_C(\omega_{+,y}) + \mu_C(\omega_{-,y}). \end{aligned} \tag{13}$$

These equations suggest reading μ_E and μ_F as *surface probabilities* that can be derived from underlying *hidden* probabilities μ_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 2 for an illustration).

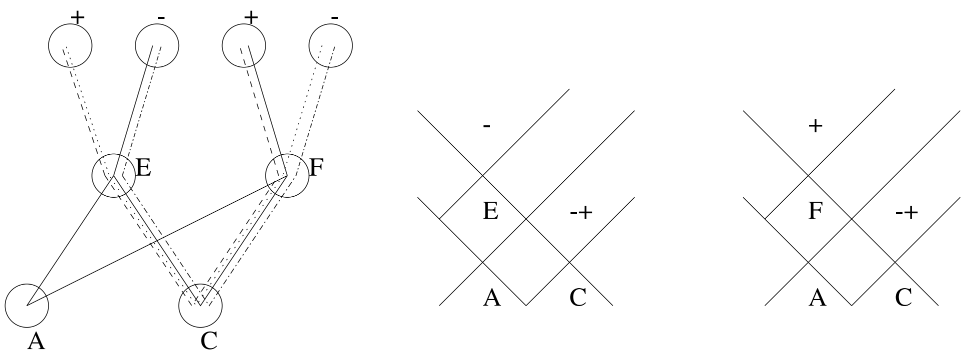

Figure 2: Determinate-value ('fatalist') hidden variable model for two upward incompatible events E and F. Left: A SOBST graph shows the causal relations between the event A of selecting either E or F, the ('hidden') event C providing the instruction sets (indicated by four different line types), the 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 'fatalist 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 '$+$'.

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 2). In this picture, 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. Thus, there is no genuine splitting (no 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 eq. (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.

Nothing in this argument depends on the fact that E and F have two atomic outcomes each. 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 fatalist construction.

We see that in the SOBST framework, the mathematical existence of the joint measure $p_{E,F}$ is interpreted as the existence of a *determinate-value hidden variable model* underlying the surface probabilities μ_E and μ_F. This model is subject to *counterfactual definiteness*: In the h.v. model, each hidden state (i.e., atomic outcome $\omega_{x,y}$ of C) predetermines the outcomes not only 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).

This brings us to the modal aspect of the situation. Initially, by specifying in SOBST two 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 h.v. 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.

Mathematically, it was trivially possible to define (classical) joint probabilities $p_{E,F}$ for pairs of outcomes of the incompatible events E and F. In SOBST, we know that (classical) probability spaces are in one-to-one correspondence with events, representing just one level of modality: which outcome will occur? Thus, the mathematical definability of the joint probability also seems to affect a collapse of the two initial levels of modality into one. This reading is confirmed by the SOBST interpretation of the joint probability as the probability of hidden states satisfying counterfactual definiteness, and it will play a crucial role in the interpretation of Fine's theorem.

5.3 Empirical implications

So far there is nothing wrong empirically with the mathematical existence of a joint probability measure for 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 QM-incompatible 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. 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.[26] 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.[27] Thus, we have the following preliminary 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.

6 Fine's theorem in SOBST models for Bell/Aspect

Section 5 has shown that for some setups one may define joint probability distributions for non-commuting observables, and there is always a SOBST interpretation for joint probabilities of outcomes of upward incompatible events. We will now show that a SOBST construction, if applied to more complicated setups, will lead to a conflict with observed data.

Thus, we will consider two SOBST models for the Bell/Aspect experiment and indicate in which way Fine's theorem applies. We will be able to follow Fine's equivalence proofs everywhere: A SOBST model allows for the definition of the required joint probability measure if and only if the Bell/CH inequalities hold in the model. On the one hand, this is not surprising, since Fine's mathematics have not been questioned. On the other hand, however, the possibility of deriving the theorem in the rich framework of SOBST allows for some interesting insights into the nature of the theorem and its implications.

6.1 Model 1: No hidden variables

Figure 3 shows a surface model of the Bell/Aspect experiment that one would posit if one had just read how to set up the experiment. No hidden states whatsoever are assumed. The model is empirically adequate: the probabilities represent empirically established values. (Among other things, this means that the probabilities in the model violate the Bell/CH inequalities.) This model is deemed explanationally inadequate. Recall the demands on SOBST graphs: The Locality Requirement is satisfied, and it can in fact be tested empirically. Screening-off, however, is violated, since there are joints of space-like events in which there are correlations. (In figure 3, one such case is indicated by thick lines.) This is what sends people looking for a common cause for

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

[27]This should already make us somewhat skeptical of Fine's (jd) condition (Fine, 1982b) for the existence of joint distributions. In the case just considered, (jd) *is* violated, but there is no empirical penalty for this. We believe that (jd) is too strong: many joint probabilities not allowed by (jd) may peacefully be defined as 'purely mathematical probabilities'. Cf. section 7.5 below for a discussion.

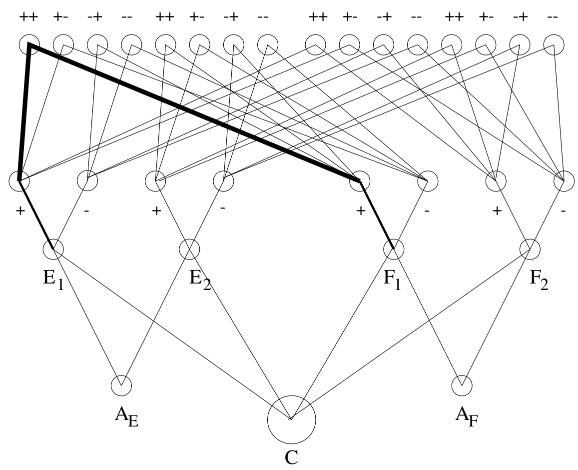

Figure 3: SOBST model of the Bell/Aspect experiment without hidden variables. C is the event of creating a particle pair, A_E (A_F) is the event of selecting the polarizer setting in the left (right) wing. E_i and F_j are the respective events of measuring the spin projection in the selected direction. 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.

the correlations, or equivalently, for hidden variables, motivating the construction of the model that we will analyze in section 6.2.

We have already remarked that the model of figure 3 does not satisfy the Bell/CH inequalities. Just as a short check, let us see why, in line with Fine's theorem, a joint probability measure for the outcomes of E_1, E_2, F_1, and F_2 cannot be defined mathematically either. In the model, there are four 'double' probabilities $\mu_{E_i \cup F_j}$, $i, j = 1, 2$. Also, we have four 'single' probabilities on outcomes μ_{E_i} and μ_{F_j}, $i, j = 1, 2$. In order to be able to talk about those 'single' probabilities consistently, the Locality Requirement of section 2.2 had to be invoked, since it guarantees for $i, j, k = 1, 2$ and outcomes e of E_i and f of F_j,

$$\begin{aligned} \mu_{E_i}(e) = \mu_{E_i \cup F_j}(e \cap \mathbf{1}_F) = \mu_{E_i \cup F_k}(e \cap \mathbf{1}_F), \\ \mu_{F_j}(f) = \mu_{E_i \cup F_j}(\mathbf{1}_E \cap f) = \mu_{E_k \cup F_j}(\mathbf{1}_E \cap f). \end{aligned} \tag{14}$$

The purported measure p_{E_1,E_2,F_1,F_2} should return all 'double' (and thus, *a fortiori*, all single) measures as marginals. This means, e.g.,

$$p_{E_1,E_2,F_1,F_2}(e, \mathbf{1}_{E_2}, f, \mathbf{1}_{F_2}) = \mu_{E_1 \cup F_1}(e \cap f). \tag{15}$$

Thus, there are two candidates for defining p_{E_1,E_2,F_1,F_2} from the given 'double' probabilities, which have to coincide for p to return all the right marginals:

$$p_{E_1,E_2,F_1,F_2}(e_1, e_2, f_1, f_2) = \mu_{E_1 \cup F_1}(e_1 \cap f_1) \cdot \mu_{E_2 \cup F_2}(e_2 \cap f_2) \tag{16}$$

$$= \mu_{E_1 \cup F_2}(e_1 \cap f_2) \cdot \mu_{E_2 \cup F_1}(e_2 \cap f_1). \tag{17}$$

This equality, however, is violated: try calculating

$$
\begin{aligned}
p_{E_1,E_2,F_1,F_2}(e_1, \mathbf{1}_{E_2}, \mathbf{1}_{F_1}, f_2) &= \mu_{E_1\cup F_1}(e_1 \cap \mathbf{1}_{F_1}) \cdot \mu_{E_2\cup F_2}(\mathbf{1}_{E_2} \cap f_2) \\
&= \mu_{E_1}(e_1) \cdot \mu_{F_2}(f_2) \qquad \text{(via (16))} \\
&= \mu_{E_1\cup F_2}(e_1 \cap f_2) \cdot \mu_{E_2\cup F_1}(\mathbf{1}_{E_2} \cap \mathbf{1}_{F_1}) \\
&= \mu_{E_1\cup F_2}(e_1 \cap f_2) \qquad \text{(via (17)).}
\end{aligned}
\tag{18}
$$

If this held, Screening-off would hold, and there would be no correlations, but that is not what the experiment (or the quantum mechanical prediction) shows. So we see that the simple, empirically adequate but explanatorily wanting SOBST model of figure 3 does not allow for the definition of the joint probability measure p_{E_1,E_2,F_1,F_2}. The breakdown of the attempted definition can be traced back to a violation of Screening-off.

If one wishes to be able to define the measure p_{E_1,E_2,F_1,F_2}, one should extend the model in such a way that Screening-off becomes satisfied. However, it is not the wish to define p_{E_1,E_2,F_1,F_2} that drives people to extending the model — it is rather the quest for explanation, guided by physical intuition, that gives the real motivation for Screening-off and thus the force behind extending the SOBST model of figure 3 by hidden variables.

6.2 Model 2: Factorizable stochastic hidden variables

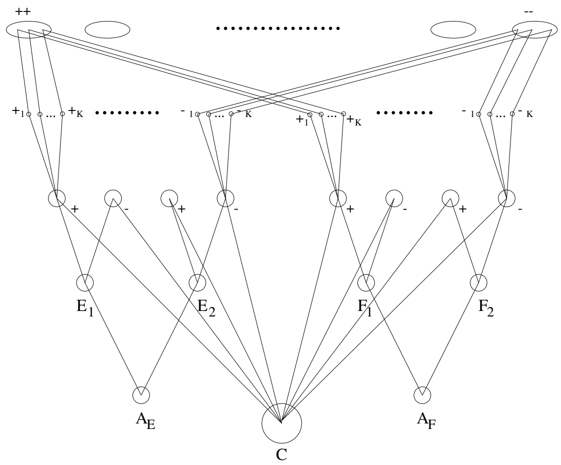

Figure 4: SOBST model with stochastic hidden variables. The K-multiplied extended outcomes are only shown for the outcome '$++$' of $E_1 \cup F_1$ and for outcome '$--$' of $E_2 \cup F_2$ in order to keep the picture tractable. Ovals on the top correspond to surface outcomes, i.e., they integrate outcomes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

Figure 4 represents a factorizable stochastic hidden variable model of the Bell/Aspect experiment. In this model, the Bell/CH inequalities are derivable.[28] Thus, while the

[28]For a proof, cf. (Placek, 1999).

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 how all the features used to specify the model play a crucial role in the definition.

The model of figure 4 includes a common common cause event C with K atomic outcomes. The extended outcomes of the events E_i and F_j are screened off conditional on C's outcomes.[29] The model is demanded to reproduce the surface probabilities of the experiment, i.e., to satisfy the condition of Adequacy. Furthermore, the model satisfies the ('hidden') Locality Requirement in the form of what is generally called 'Locality' and 'No Conspiracy'. Based on this model, we define the joint probability measure[30] as

$$\begin{aligned} p_{E_1,E_2,F_1,F_2}(e_1, e_2, f_1, f_2) = & \\ \sum_{i=1}^{K} & \overset{*}{\mu}_{E_1\cup C}\, (e_1 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{E_2\cup C}\, (e_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \\ & \overset{*}{\mu}_{F_1\cup C}\, (f_1 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{F_2\cup C}\, (f_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{C}\, (\omega_i), \end{aligned} \tag{19}$$

where $\overset{*}{\mu}$ indicates probability measures defined in the extended model. In this definition, we made use of the 'hidden' Locality Requirement in order to define, e.g.,

$$\overset{*}{\mu}_{E_1\cup C}\, (e_i \cap \omega_j) = \overset{*}{\mu}_{E_1\cup F_1\cup C}\, (e_i \cap \mathbf{1}_{F_1} \cap \omega_j) = \overset{*}{\mu}_{E_1\cup F_2\cup C}\, (e_i \cap \mathbf{1}_{F_2} \cap \omega_j), \tag{20}$$

and (the condition of 'No Conspiracy')

$$\overset{*}{\mu}_{C}\, (\omega_k) = \overset{*}{\mu}_{E_i\cup F_j\cup C}\, (\mathbf{1}_{E_i} \cap \mathbf{1}_{F_j} \cap \omega_k). \tag{21}$$

Appendix B shows that this measure 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 which also allowed for the derivation of the Bell/CH inequalities. If one of the conditions is dropped, the joint measure is not consistently definable.

In addition to returning the 'double' joint surface probabilities for 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,

$$\begin{aligned} p_{E_1,E_2}(e_1, e_2) = p_{E_1,E_2,F_1,F_2}(e_1, e_2, \mathbf{1}_{F_1}, \mathbf{1}_{F_2}) = & \\ \sum_{i=1}^{K} \overset{*}{\mu}_{E_1\cup C}\, (e_1 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{E_2\cup C}\, (f_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{C}\, (\omega_i). & \end{aligned} \tag{22}$$

Note that the above joint probability of eq. (19) is *not* interpreted as the probability of an outcome of some event (i.e., as the probability of a hidden state) — in the

[29]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.

[30]Cf. (Fine, 1982a, eq. (3)).

framework of a factorizable stochastic model, the above joint probability is rather a weighed average over the probabilities of hidden states. Nevertheless, as we will see in section 7, whenever the above joint probability p_{E_1,E_2,F_1,F_2} is definable in a factorizable stochastic model, we may as well construct a fatalist model, with probabilities assigned to outcomes of a (hidden) common cause event.

6.3 Comparison of the two models

The two SOBST models considered above show that trying to model the space-time and modal structure of the Bell/Aspect experiment leads to a dilemma: Either the model is empirically adequate, but explanatorily wanting (model 1), or the model is empirically inadequate (model 2).

Model 1 correctly reproduces the experimental data. However, it is not fine-grained enough to explain all the correlations between space-like events present in the data (cf. again the thick line in figure 3).

Faced with this difficulty, one option is to allow for such unexplained correlations — maybe backed by the argument that in science, the notion of what needs to be explained is as much subject to historical development as are scientific theories.[31] While this might be a good move to make in the end, we will not explore it any further here.

Model 2 follows a different strategy. In model 2, additional 'hidden' structure (a common common cause event C) is added in order to make the model explanatorily satisfying. In addition to adding event C, model 2 is subject to a number of constraints that are selected on the basis of physical intuition. The condition of Adequacy simply makes sure that model 2 is a model of the same surface data as model 1; it is therefore the most basic condition. The condition of Screening-off embodies a conviction about how correlations can be (causally) explained. Finally, the Locality Requirement makes sure that the distant correlations that model 1 left unexplained are really explained in a local way, and it also rules out 'conspiratorial' influences from the direction setting in the two wings on the particle source.

The predictions of QM, model 1, and also the Bell/Aspect experiment itself yield probabilities that violate the Bell/CH inequalities. Thus, model 2 must not be constrained by these inequalities — otherwise, it will be empirically inadequate. In the light of Fine's theorem this means that model 2 should be constructed in such a way as not to put all the data into one large probability space. Consequently, model 2 works with many small probability spaces.

However, the discussion of section 6.2 shows that this strategy for constructiong model 2 is doomed to fail: in model 2, the Bell/CH inequalities are derivable, and a joint probability distribution for all the quantum observables involved can be constructed. So despite the caution of working with separated probability spaces only, the conditions of Adequacy, Screening-off, and the Locality Requirement allowed one to glue the separated spaces together to one large joint space. — Section 8 will explore the option of working with generalized probability theory in order to circumvent Fine's theorem.

[31] Cf. (Van Fraassen, 1989; Fine, 1989). — Of course, these authors opted for this solution only after exploring the other options.

7 The interpretation of Fine's theorem

In section 6.2 we showed a SOBST model for the Bell/Aspect experiment that allowed for the definition of a joint probability measure p_{E_1,E_2,F_1,F_2}. Here, we will comment on the implications of this derivation.

7.1 The interpretation of joint probabilities for incompatible events

In section 5 we already touched upon the issue of the SOBST interpretation of probabilities like p_{E_1,E_2} of equation (22). The upshot was that while a conjunctive reading of such probabilities was impossible, they could be interpreted as probabilities on determinate-value hidden variables. This reading is thus also possible for p_{E_1,E_2}.

More generally, and again in line with Fine's equivalences, we can give a determinate-value hidden variable reading to the whole probability measure p_{E_1,E_2,F_1,F_2}. First note that in specifying the model of figure 4, we have only demanded stochastic hidden variables satisfying the condition of Screening-off, not determinate-value hidden variables. Thus, the number K of atomic outcomes of the common common cause C does not have to coincide with the number of possible 'instruction sets' in a determinate-value h.v. model. This shows that we cannot simply read the probabilities p_{E_1,E_2,F_1,F_2} as probabilities on the outcomes of C in the given model. However, preserving the location of C in the model, we can 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, and in which these atomic outcomes are assigned exactly the probabilities p_{E_1,E_2,F_1,F_2}. Thus, the mathematical definability of the joint measure p_{E_1,E_2,F_1,F_2} subject to the mentioned constraints (i.e., returning given joint surface probabilities as marginals) 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:

$$p_{E_1,E_2,F_1,F_2}(e_1, e_2, f_1, f_2) = \mu_C(\omega_{e_1,e_2,f_1,f_2}). \tag{23}$$

In this way, we have given a SOBST interpretation of one more of the equivalences Fine has established:[32] *The probability measure p_{E_1,E_2,F_1,F_2} can be defined to return the given 'double' probabilities as marginals if and only if we can give a factorizable determinate-value hidden variable model for the situation.*

7.2 Determinate-value vs. stochastic hidden variables

There has been some debate as to whether stochastic h.v. models are more general than determinate-value ones. Fine and Suppes/Zanotti have offered equivalence proofs to the effect that stochastic models are in fact not more general than determinate-value models: Given surface probabilities admit a stochastic model if and only if they admit a determinate-value model.

Fine's argument is basically the discussion of section 7.1, which showed that from a stochastic h.v. model we can construct a determinate-value model preserving the surface probabilities, together with the observation that any determinate-value h.v. model

[32](Fine, 1982a). He uses 'deterministic' for 'factorizable determinate-value'.

is also a stochastic model (with trivial probabilities). In this argument, the joint probability measure p_{E_1,E_2,F_1,F_2} plays only an auxiliary role.

Suppes and Zanotti[33] have given an argument that does not invoke a joint measure, but that draws on physical facts about the Bell/Aspect experiment: In order for a stochastic h.v. model to be able to reproduce the perfect correlations observed with parallel settings in both wings, the model has to have trivial probabilities. As at the time of pair creation, parallel settings are always a possibility (even an experimental possibility, given in-flight switching), any stochastic h.v. model used to describe the situation has to have trivial probabilities in any case — thus, it has to be a determinate-value model in the first place.

These two proofs establish that the class of surface probabilities that admit a stochastic h.v. model is no larger than the class that admits a determinate-value model. This may appear quite striking: recall that 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.

In determinate-value (fatalistic) models, the probabilities $\mu_C(\omega_i)$ of outcomes of the common cause event are generally nontrivial, but they can naturally be interpreted epistemically, thus allowing for an ignorance interpretation of the statistical character of QM — this has always been one of the chief motivations for constructing such models.

Stochastic models, on the other hand, are motivated by the feeling that the determinism embodied in fatalistic models might be too restrictive to allow for reproducing the experimental data. By allowing for two kinds of nontrivial probabilities (on the outcomes of the common cause and on the measurement results conditional on such outcomes), an epistemic reading is hardly possible any more. The demand for Screening-off itself appears not too restrictive.

Now Fine's proof shows that given a stochastic model satisfying Screening-off, a determinate-value model 'beneath' comes at no extra cost: by reiterating a hidden variable construction, you do not impose any new constraints on the surface probabilities that the model is able to reproduce.

A closer look at the situation reveals that for a proponent of stochastic hidden variables, the situation is actually worse than that. Even barring the physically motivated argument by Suppes and Zanotti to the effect that one should not consider stochastic models in the first place, 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.[34] The fatalistic construction, on the other hand, faces no such limitations. 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 *simpliciter*).

7.3 Fine's circle of equivalences

Fine has established a whole number of equivalences, not all of which have been mentioned so far. Here we give a summary in the SOBST framework.

[33](Suppes and Zanotti, 1976).

[34]For a proof, cf. (Placek, 1999).

THEOREM 4
Assume a SOBST surface model $\langle W, \leq, \mathcal{E}, \Upsilon \rangle$ *is given in which there are events* E_i *and* F_j, $i, j = 1, 2$, *each with atomic outcomes '+' and '−', such that (i)* E_1 *and* E_2 *are upward incompatible, (ii)* F_1 *and* F_2 *are upward incompatible, and (iii)* E_i *and* F_j *are upward compatible and space-like separated,* $i, j = 1, 2$. *The (surface) Locality Requirement is assumed to hold, i.e.,* $\mu_{E \cup F}(e \cap \mathbf{1}_F) = \mu_E(e)$, $\mu_{E \cup F}(\mathbf{1}_E \cap f) = \mu_F(f)$. *Then the following conditions are equivalent:*

1. *We can construct an equivalent SOBST model (i.e., one having the same surface probabilities* $\mu_{E_i \cup F_j}$, μ_{E_i}, *and* μ_{F_j}*) that contains a common common cause* C *satisfying the conditions of the Locality Requirement, No Conspiracy, Screening-off, and Adequacy, i.e., a factorizable stochastic hidden variable model.*

2. *The probabilities* $\mu_{E_i \cup F_j}$, μ_{E_i}, *and* μ_{F_j} *satisfy the following four (Bell/CH) inequalities:*

$$\begin{aligned} -1 \quad \leq \quad & \mu_{E_i \cup F_j}(+\cap+) + \mu_{E_i \cup F_{j'}}(+\cap+) + \mu_{E_{i'}, \cup F_{j'}}(+\cap+) \\ & - \mu_{E_{i'}, \cup F_j}(+\cap+) - \mu_{E_i}(+) - \mu_{F_{j'}}(+) \quad \leq \quad 0 \end{aligned} \tag{24}$$

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

3. *It is possible to define mathematically consistently joint probability measures for all pairs and triples of outcomes of compatible as well as incompatible events such that these measures return the four joint probabilities* $\mu_{E_i \cup F_j}$ *(and thus* a fortiori *the four single probabilities* μ_{E_i} *and* μ_{F_j}*) as marginals.*

4. *It is possible to define mathematically consistently a joint probability measure* p_{E_1, E_2, F_1, F_2} *that returns the four joint probabilities* $\mu_{E_i \cup F_j}$ *(and thus* a fortiori *the four single probabilities* μ_{E_i} *and* μ_{F_j}*) as marginals.*

5. *It is possible to give an equivalent SOBST model (i.e., one having the same surface probabilities* $\mu_{E_i \cup F_j}$, μ_{E_i}, *and* μ_{F_j}*) that contains an event* C *whose 16 atomic outcomes act as determinate-value hidden variables.*

6. *Pitowsky's condition:*[35] *The vector of surface probabilities* $p = \langle \mu_{E_1}(+), \mu_{E_2}(+), \mu_{F_1}(+), \mu_{F_2}(+), \mu_{E_1 \cup F_1}(+\cap+), \mu_{E_1 \cup F_2}(+\cap+), \mu_{E_2 \cup F_1}(+\cap+), \mu_{E_2 \cup F_2}(+\cap+) \rangle$ *is an element of the Clauser-Horne-polytope* $c(4, S)$ *with* $S = \{\{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}\}$.

For the implication $1 \to 2$, cf. (Placek, 1999). $1 \to 4$ has been shown in section 6.2, while $4 \leftrightarrow 5$ was the subject of section 7.1. $5 \to 1$ is trivial.[36] $6 \leftrightarrow 4$ was treated in section 3.4. For $4 \leftrightarrow 3$ as well as $2 \leftrightarrow 3$, which are both simple mathematical equivalences, we refer to Fine.[37]

7.4 Are the Bell/CH inequalities to be expected?

For the full Bell/Aspect experiment, i.e., for two possible settings in both wings, the existence of a joint probability measure has been shown above to depend crucially on

[35] Cf. (Pitowsky, 1989, Chap. 2.5, pp. 27ff.). See also section 3.3 above.

[36] Recall from section 7.2 that in accordance with standard usage, we take 'stochastic' here in a broad sense, i.e., including trivial probabilities.

[37] (Fine, 1982a).

details of the underlying model. The model must be quite ‘rich’, i.e., satisfy quite a number of constraints, for p_{E_1,E_2,F_1,F_2} to be definable. We take it that it is rather this rich underlying structure of the SOBST model than the mere mathematical existence of a joint probability measure which allows for an interpretation of the empirical violation of the Bell/CH inequalities: the Bell/Aspect experiment shows that a certain ideal of explanation that requires (among other things) Screening-off for the observed correlations cannot be satisfied.

The definability of the joint probability measure p_{E_1,E_2,F_1,F_2} does not speak against this interpretation, as we have just remarked. Note that this troublesome measure can also simply be viewed as an auxiliary step in the construction of a determinate-value hidden variable model.

Actually, in line with Fine’s mathematics, but not with his remarks, we advocate the following logical identity thesis: in SOBST models, the definability of p_{E_1,E_2,F_1,F_2} is mathematically the same as the satisfaction of the Bell/CH inequalities. This means that starting from either of these conditions, we can enter the search for the structure of SOBST models that causes the condition to hold. Thus, Fine’s theorem does not speak against the interpretation that the violation of the Bell/CH inequalities proves something about possible explanations of the Bell/Aspect experiment.

7.5 When is asking for joint probabilities asking for trouble?

We have seen that in the Bell/Aspect experiment, asking for the existence of the joint probability measure p_{E_1,E_2,F_1,F_2} *is* asking for trouble: models in which this measure is consistently definable will allow the derivation of the empirically violated Bell/CH inequalities. By the remarks about the circle of equivalences, we see that this holds both ways.

However, just asking for joint probabilities for outcomes of incompatible events (e.g., quantum-mechanically: for results of measurements of non-commuting observables) is not generally enough to cause trouble. As we have remarked in section 5.3 above, for any pair of such events, joint probabilities are always and uncontroversially definable. Furthermore, in the Bell/Aspect experiment, if the setting is fixed in one wing, it is possible to assign consistent joint probabilities as well.[38]

Fine has coined his condition (jd) in such a way that the following equivalence holds: *Given a set of quantum observables, all pairs commute if and only if the set satisfies (jd).*[39] Thus, (jd) does indeed offer some insights on the relation between ‘quantum mechanical wisdom’ and compatible vs. incompatible events. However, the equivalence means that (jd) is too strong to single out joint probabilities that cause trouble on the level of surface probabilities.

It remains interesting to ask which demands for joint probability measures do have detrimental effects, i.e., empirically testable consequences in terms of surface probabilities deviating from the quantum mechanical predictions. ‘Quantum mechanical wisdom’ alone will be a guide as to which measurement events can be assumed to be compatible, but this alone does not answer the question. We seem to need a thorough analysis of the experiment to be modeled in every case. At any rate, we have not been able to derive a general theorem along these lines. Still, it is clear that Fine’s theorem linking the existence of the joint measure p_{E_1,E_2,F_1,F_2} to the Bell/CH inequalities will be a prominent special case.

[38](Landau, 1987).

[39](Fine, 1982b).

8 Is generalized probability of any help?

The discussion of section 6.3 showed that trying to model the Bell/Aspect experiment, one seems to face a dilemma: either opt for a model that gets the probabilities right, but that does not explain distant correlations (model 1: quantum state), or give a model that explains the correlations, but gets the probabilities all wrong (model 2: equivalent to the existence of a classical state / fatalist model). Fine's theorem showed that, without intending this, model 2 allowed for a 'large' joint probability distribution, the existence of which was seen to be equivalent to the satisfaction of the Bell/CH inequalities. Maybe if one could state one's intentions clearer, one might be able to give a better model?

The idea is the following: in specifying model 2, physical intuition was the guide to selecting the model structure. It turned out that the model thus constructed contained unnecessary structure, namely, the (purely mathematical) joint probabilities like p_{E_1,E_2} that would determine outcomes of impossible measurements. These purely mathematical entities were seen to have a detrimental effect on the model's empirical adequacy: from the bogus structure of model 2, the Bell/CH inequalities were derivable. So it might be worth while to let structure be our guide, by specifying a model in such a way that no unnecessary information builds up. This means that, e.g., joint probabilities for non-commuting observables should not even be defined in the model. We have seen that on the basis of classical probability theory, this end could not be achived. It therefore seems natural to explore whether some probability structures weaker than classical probability spaces could help. Note that one very natural class of probability structures, namely the quantum state itself, which leads to empirically correct results, has been ruled out already as not explanatorily satisfying. It is therefore necessary to investigate some stronger structures. We will concentrate on so-called generalized probability theory.

8.1 The generalized probability

In classical probability theory, the main algebraic structure is a σ-field of subsets of a given set, i.e., a Boolean algebra. Generalized probability theory starts with an algebraically weaker structure.

A family C of subsets of a non-empty set Ω is a *σ-additive class* iff it satisfies (Gudder, 1988):

1. $\Omega \in C$;
2. if $a \in C$, then $a' \in C$;
3. if $a_1, a_2, \ldots \in C$ are mutually disjoint, then $\bigcup_i a_i \in C$,

where a' denotes the complement of a. Note this feature of the third clause above: to exclude $a_1 \cup a_2$ from C, a_1 and a_2 must not be disjoint; however, if we only wish, we may include $a_1 \cup a_2$ in C, even if a_1 and a_2 are not disjoint.

A (generalized) probability measure on a σ-additive class C of subsets of Ω is a function $\mu : C \to [0, 1]$ such that

1. $\mu(\Omega) = 1$;
2. if $a_1, a_2, \ldots \in C$ are mutually disjoint, then $\mu(\bigcup_i a_i) = \sum_i \mu(a_i)$.

A generalized probability space is a triple $\langle \Omega, C, \mu \rangle$, where Ω is a non-empty set, C is a σ-additive class of subsets of Ω, and μ is a generalized probability measure on C (see (Gudder, 1969)).

The minimalist interpretation puts the blame for the derivation of the Bell inequalities on definability of joint probabilities for results of non-commuting observables. Definability of these probabilities, however, is a natural by-product of working with the classical probabilities. The generalized probability introduced above permits to circumvent the problem, that is, to leave undefined the probability of joint occurrence of results of non-commuting observables.

The basic trick is this: Suppose we are going to model results A_1+ and A_2+ of spin measurements with non-parallel directions A_1 and A_2. Obviously, we need A_1+ and A_2+ to be 'events', that is, elements of σ-additive class C on some set Ω. We however object to probability being assigned to the 'simultaneous occurrence' $A_1+ \cap A_2+$. Thus, $A_1+ \cap A_2+ = (A_1+' \cup A_2+')'$ must not be in C, and it can only be so, given the definition of σ-additive class, if A_1+' and A_2+' are not disjoint, i.e,

$$A_1+' \cap A_2+' \neq \emptyset.$$

To explain now more precisely what it means that some data admits a hidden variable model with the generalized probability, let the vector

$$\langle p_1, \dots p_n, \dots p_{ij} \dots \rangle,$$

with $\langle i, j \rangle \in S = \{\langle k, l \rangle : 1 \leq k < l \leq n\}$ and all $p_i, p_{ij} \in [0, 1]$ represent probabilistic data collected in some experiments. The vector of this kind is called the probabilistic data of (n, S)-type.

DEFINITION 11
We say that the data of (n, S)-type admits a hidden variable model with the generalized probability if and only if there is a generalized probability space $\langle \Omega, C, \mu \rangle$ and sets $A_1, \dots A_n \in C$ such that

$$p_i = \mu(A_i) \text{ and } p_{ij} = \mu(A_i \cap A_j)$$

for all $i = 1, \dots, n$ and $\langle i, j \rangle \in S$.

We will now show below that working with the generalized probabilities allows one to construct a determinate-value h.v. model for the Bell-Aspect experiment that is empirically adequate, that is, it is *not* committed to the Bell inequalities. To this end we will explicitly construct the model.

8.2 The Model of Bell-Aspect with Generalized Probabilities

To recall the experimental setup, if A_i and B_i $(i = 1, 2)$ are measurements of spin projections in the left and right wing, respectively, the angles between polarization directions are these:

$$\angle(A_1, B_1) = 120^\circ \quad \angle(A_1, B_2) = 240^\circ \quad \angle(A_2, B_1) = 0^\circ \quad \angle(A_2, B_2) = 120^\circ.$$

Most generally, one may consider the following angles:

$$\angle(A_1, B_1) = \alpha \quad \angle(A_1, B_2) = \beta \quad \angle(A_2, B_1) = \gamma - \alpha \quad \angle(A_2, B_2) = \gamma - \beta.$$

The A_i observables are non-commuting as are the B_i observables. As usual, we will refer to results of the corresponding measurements by A_i+, A_i-, B_i+, and B_i-.

We need now to assign the results to some subsets of a set Ω. To simplify, $A_+, \ldots$ refer henceforth to the subsets. We need to do this assignment in a way that $A_i \pm \cap A_j \pm$ and $B_i \pm \cap B_j \pm$ are *not* in the σ-additive class, whereas $A_i \pm \cap B_j \pm$ are. This translates, as showed above, into this condition:

$A_i \pm' \cap A_j \pm' \neq \emptyset$ and $B_i \pm' \cap B_j \pm' \neq \emptyset$.

The simplest model that satisfies this condition and respects the QM probabilities is this. As the master set, take any 12-elements set Ω, e.g.,

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, x, y\}.$$

Now, the results are assigned the following subsets of Ω:

$$\begin{gathered} A_1+ = \{1, 4, 6, 7, 8, x, y\} \;\; A_1- = \{0, 2, 3, 5, 9\} \;\; A_2+ = \{1, 2, 3, 4, 5, x\} \\ A_2- = \{0, 6, 7, 8, 9, y\} \;\; B_1+ = \{5, 6, 7, 8, 9, x\} \;\; B_1- = \{0, 1, 2, 3, 4, y\} \\ B_2+ = \{1, 2, 3, 6, 9, x, y\} \;\; B_2- = \{0, 4, 5, 7, 8\} \\ . \end{gathered}$$

All these subsets will be assigned measure $1/2$. Note that the subsets assigned to $-$ results are complement of the subsets assigned to the corresponding $+$ results. The trick of this construction is to find such subsets of Ω that the intersections $A_i \pm \cap B_j \pm$ be the atoms of the resulting additive class. It remains to display these intersections, together with the measures assigned.

With these stipulations, the σ-additive class C of subsets of Ω comprises, apart from Ω itself and $\emptyset$, the subsets of only these six kinds:

$$A_i\pm, A_i\pm', B_i\pm, B_i\pm', A_i\pm' \cup B_i\pm', (A_i\pm' \cup B_i\pm')' = A_i\pm \cap B_i\pm \,.$$

It remains to explicitly display the subsets assigned to $A_i \pm \cap B_i \pm$ and the probabilities assigned to by QM algorithm. (The number to the right of a subset represents the probability assigned.)

Result	Set	Specific measure	QM measure, generally
$A_1+ \cap B_1+$	$\{6,7,8,x\}$	$3/8$	$1/2 \sin^2 \frac{\alpha}{2}$
$A_1+ \cap B_2+$	$\{1,6,x,y\}$	$3/8$	$1/2 \sin^2 \frac{\beta}{2}$
$A_2+ \cap B_1+$	$\{5,x\}$	0	$1/2 \sin^2 \frac{\gamma-\alpha}{2}$
$A_2+ \cap B_2+$	$\{1,2,3,x\}$	$3/8$	$1/2 \sin^2 \frac{\gamma-\beta}{2}$
$A_1- \cap B_1-$	$\{0,2,3\}$	$3/8$	$1/2 \sin^2 \frac{\alpha}{2}$
$A_1- \cap B_2-$	$\{0,5\}$	$3/8$	$1/2 \sin^2 \frac{\beta}{2}$
$A_2- \cap B_1-$	$\{0,y\}$	0	$1/2 \sin^2 \frac{\gamma-\alpha}{2}$
$A_2- \cap B_2-$	$\{0,7,8\}$	$3/8$	$1/2 \sin^2 \frac{\gamma-\beta}{2}$
$A_1- \cap B_1+$	$\{5,9\}$	$1/8$	$1/2 \cos^2 \frac{\alpha}{2}$
$A_1- \cap B_2+$	$\{2,3,9\}$	$1/8$	$1/2 \cos^2 \frac{\beta}{2}$
$A_2- \cap B_1+$	$\{6,7,8,9\}$	$1/2$	$1/2 \cos^2 \frac{\gamma-\alpha}{2}$
$A_2- \cap B_2+$	$\{6,9,y\}$	$1/8$	$1/2 \cos^2 \frac{\gamma-\beta}{2}$
$A_1+ \cap B_1-$	$\{1,4,y\}$	$1/8$	$1/2 \cos^2 \frac{\alpha}{2}$
$A_1+ \cap B_2-$	$\{4,7,8\}$	$1/8$	$1/2 \cos^2 \frac{\beta}{2}$
$A_2+ \cap B_1-$	$\{1,2,3,4\}$	$1/2$	$1/2 \cos^2 \frac{\gamma-\alpha}{2}$
$A_2+ \cap B_2-$	$\{4,5\}$	$1/8$	$1/2 \cos^2 \frac{\gamma-\beta}{2}$

Finally, as already mentioned, all subsets $A_i\pm$ and $B_i\pm$ are assigned measure $1/2$.

Although this probability assignment may look odd for our classically trained intuitions, nevertheless, it is mathematically correct. Moreover, it is straightforward to show that the model is applicable to the Bell-Aspect setup with any polarization directions, and hence, with other probability assignments (sse the last column above).

Thus, the minimalist interpretation has a success: there is a determinate-value h.v. model for the Bell-Aspect experiment that is empirically adequate and hence not committed to the Bell inequalities; the (non-classical) probability calculus applied in this model is such that no probability is assigned to joint occurrences of results of non-commuting observables.

As we will shortly see, however, models with generalized probabilities are in conflict with some other QM predictions.

8.3 Generalized probability and Gleason's theorem

Following (Malament, 1996) we are going to show in this section that for sufficiently rich data, the hidden variable models with the underlying generalized probability, run in conflict with Gleason's theorem. Let us first introduce the notion of *quantum data set*.

DEFINITION 12
A vector $\langle p_1, \dots, p_n, \dots p_{ij} \dots \rangle$ of real numbers from $[0,1]$ and of type (n,S) is a quantum data set *if there is a Hilbert space $\mathcal{H}$, a set $\mathcal{E} = \{E_1, E_2, \dots, E_n\}$ of projection operators on $\mathcal{H}$, and a density operator W on $\mathcal{H}$ such that:*

1. *for any $E \in \mathcal{E}$, $E^{\perp} \in \mathcal{E}$ as well;*
2. *for any orthogonal E_i and E_j from $\mathcal{E}$, the span $E_i \vee E_j$ is in $\mathcal{E}$ as well;*
3. *for all i and j ($1 \leq i < j \leq n$, $\langle i,j \rangle \in S$ iff E_i and E_j are compatible (i.e., $E_iE_j = E_jE_i$.)*

4. *for all* i, $p_i = Tr(WE_i)$ *and for all* $\langle i,j\rangle \in S$, $p_{ij} = Tr(WE_iE_j)$.

THEOREM 5 (MALAMENT)
For quantum data set (QDS) $\langle p_1, \ldots, p_n, \ldots p_{ij} \ldots\rangle$ *of type* (n, S) *and corresponding* $\mathcal{H}$, $\mathcal{E}$, *and* W, *if the data set admits a hidden variable theory with the underlying generalized probability, then there exists a map* $\varphi : \mathcal{E} \to \{0,1\}$ *such that*

1. $\varphi(I) = 1$, *where* I *is the identity projection on* $\mathcal{H}$*;*

2. *for all orthogonal* E_i *and* E_j *from* $\mathcal{E}$, $\varphi(E_i \vee E_j) = \varphi(E_i) + \varphi(E_j)$.

SKETCH OF PROOF
Any generalized probability model (GPM) for QDS must assign subsets of some set Ω to all the projectors from $\mathcal{E}$, say $E \longrightarrow A_E$, so that the generalized probability measure μ match the quantum probabilities p_i, that is

$$p_i = Tr(WE_i) = \mu(A_{E_i})$$

¿From the outset we may require that $I \longrightarrow \Omega$. We will first argue that if there is a GPM for QDS, then there is a (possibly different) GPM model for the same data such that for any orthogonal projectors E and F, (1) A_E and A_F are disjoint and (2) the set $A_{E\vee F}$ assigned to the span $E \vee F$ satisfies: $A_{E\vee F} = A_E \cup A_F$

To begin with task (1), if only $p_E \neq 0$, it cannot be that $A_E \subseteq A_F$, since otherwise $0 \neq p_E = \mu(A_E) = \mu(A_E \cap A_F) = p_{EF} = 0$. But, if $p_E = 0$ for some E, and in the initial model A_E is non-empty, we may, without destroying probability relations, assign to E the empty set rather than the non-empty. Secondly, if in one model we have overlapping sets A_E and A_F corresponding to orthogonal projectors E and F, then it must be that $\mu(A_E \cap A_F) = 0$, and this permits the following 'disjoining' operation $*$ to be performed without destroying the involved probabilities:

For any pair of orthogonal projectors E and F, put: $E^* = E/(E \cap F)$, $F^* = F/(E \cap F)$ and $\mu(E^*) = \mu(E)$, $\mu(F^*) = \mu(F)$. This operation yields a correct model for QDS, if the initial model was correct.

Turning now to task (2), the initial GPM model can be uneconomical in the sense that, for orthogonal E and F, apart from containing $A_E, A_F, A'_E, A'_F, A_E \cup A_F, A_E \cap A_F$ etc., it contains an extra subset $A_{E\vee F}$ such that $A_{E\vee F} \neq A_E \cup A_F$. If this is so, we may create a simpler model by first removing this $A_{E\vee F}$ from the class and making $A_{E\vee F} = A_E \cup A_F$. The axioms of the generalized probability will guarantee that $p_{E\vee F} = p_E + p_F$, as desired.

Thus, if there is a GPM model for QSD, there is a model with orthogonal projectors represented by disjoint subsets of Ω. Let us now form from the σ-additive class C that underlies this model, the sigma-algebra $\Sigma(C)$ generated by elements of C. Obviously, Ω is the unit element of this algebra. But any Boolean algebra can be mapped into $\{0,1\}$ such that:

1. $\phi(\Omega) = 1$;

2. for all disjoint A_E and A_F from $\sigma(C)$, $\phi(A_E \cup A_F) = \phi(A_E) + \phi(A_F)$.

The restriction of ϕ to C combined with the map $\longrightarrow$ is the sought map $\varphi : \mathcal{E} \to \{0,1\}$.
□

This result runs in conflict with Gleason's theorem (Gleason, 1957), or rather, a consequence of it known as the Kochen-Specker paradox (Kochen and Specker, 1967).

Gleason's theorem says that for any Hilbert space $\mathcal{H}$ of the dimension greater than 2, every probability measure on the lattice of projectors on $\mathcal{H}$ is given by QM algorithm, for some state W. An immediate corollary of this theorem is the nonexistence of $\{0,1\}$-measures (known as dispersion-free states) on the lattice of projectors on $\mathcal{H}$, if $dim(\mathcal{H}) > 2$. Of course, this does not mean that for a finite number of projectors on $\mathcal{H}$ it is impossible to find a dispersion-free state, i.e., a function φ from this set of projectors to $\{0,1\}$ that satisfies:

$\varphi(I) = 1$ and for all orthogonal E_i and E_j, $\varphi(E_i \vee E_j) = \varphi(E_i) + \varphi(E_j)$.

However, a remarkable fact due to Kochen and Specker is that for a certain finite set of projectors on $\mathcal{H}^3$ (they used 117 projectors) no such function can exist. Thus, we must expect that the hidden variable model with the generalized probability will come in conflict with QM probabilistic predictions if the setup of the experiment is sufficiently rich, i.e., its description uses the projectors of Kochen-Specker example. The purported existence of a determinate-value h.v. model based on the generalized probability implies the existence of a map of the kind mentioned above, from the totality of the projectors involved to the $\{0,1\}$ set, and this is exactly the map prohibitted by the Kochen-Specker result. An experiment of this sort has been neither performed nor even proposed. Nevertheless, since the number of projectors involved is finite, though quite large, one should be able to propose such an experiment. This, if performed, would decide about the overall correctness of h. v. models based on the generalized probability. Although in experimental sciences nothing is certain before the experiment is performed, knowing about the astounding adequacy of QM, we may nevertheless be quite certain that the experiment will tell against these h.v. theories.

9 Conclusions

To get more insight into the nature of joint probabilities of quantum mechanical phenomena and in particular, to combat the minimalist interpretation of violations of Bell's inequalities, we showed above how joint probability distributions are interpreted in space-time structures, where we used to this end models of SOBST. As a preliminary move, we argued that the known arguments to the effect that joint probability distributions cannot be defined for results of observables that quantum mechanics describes as non-commuting, are incorrect. Then we constructed a SOBST interpretation of joint probabilities of outcomes of upward incompatible events, in which the joint probability is interpreted as the (single) probability of outcomes (=hidden state) of an added event, a common cause. This construction we applied, against the spirit of orthodox quantum mechanics, to non-commuting QM observables, and noted that for some setups, this does not lead to a conflict with observed probabilities. However, for more complicated setups, like that of the Bell-Aspect experiment, the construction led to the derivation of the Bell-CH inequality in the SOBST model, which after all should be expected from Fine's theorem. We showed, moreover, how all the conditions used in this derivation ensure the existence of the joint probability distribution Fine's theorem refers to. Thus, we concluded, the mathematical meaning of these condition is such that they allow the introduction of this joint probability; the mere existence of joint probabilities for results of *some* non-commuting observables does not land us in trouble.

We finished this part of the paper by giving the SOBST reading of a circle of equivalences Fine has established and then showing in our framework that stochastic factorizable models are *not* less restrictive than determinate value h.v. models.

In the final part of the paper, we considered the generalized probability approach to the Bell-Aspect experiment, as this non-classical probability allows us to leave undefined the probabilities of results of non-commuting observables. We showed, in agreement with Malament results, that in this probabilistic framework the troublesome Bell-CH inequalities are not derivable. However, the probabilities calculated by some other experiment are in conflict with probabilistic predictions of QM.

Thus, we conclude, it is not definability of joint probabilities for non-commuting observables that bears the blame for Bell's inequalities.

A Common causes in the SOBST framework

The idea of a common cause as advocated by (Reichenbach, 1956), adapted to the SOBST framework, is the following: Assume you are given a SOBST surface model that contains unexplained correlations between the outcomes of space-like separated events E and F. This model is explanatorily inadequate. The correlations are explained if you can give an adequate 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 away that in each of the partitions, the correlations vanish.)

Technically, the common cause event C will have K atomic outcomes. The results r^e and r^f of E and F will be replaced by K copies r^{ei} and r^{fj} each such that in each atomic outcome ω_k of C there will be just the corresponding results r^{ek} and r^{fk} of E and F. The outcomes of E and F are extended accordingly, so that, e.g., the extended outcome e^* of E comprises all the multiplied results r^{ei}, $i \leq K$. In the model thus extended, there is now new freedom to adjust the (hidden) probabilities $\mu^*_{E\cup F\cup C}(e \cap f \cap \omega_i)$ in such a way that conditional on the ω_i, 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*.[40] 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.[41] 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 3). Therefore, we define a common common cause as follows:[42]

DEFINITION 13 (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 model* $\langle W, \leq, \mathcal{E}, \Upsilon\rangle$. *Let* $E^n \cup F^n$ *and* $E^m \cup F^m$ *be incompatible for all* $n \neq m$ $(n, m \leq N)$ *and some event* S *exist such that* $S \prec E^n$ *and* $S \prec F^n$ *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, *which corresponds to results* $r^{e1n}, \ldots, r^{eIn}$ *and* $r^{f1n}, \ldots r^{fJn}$, *respectively. 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* $j =\leq J$:

$$\mu_{E^n \cup F^n}(e^n_i \cap f^n_j) \neq \mu_{E^n\cup F^n}(e^n_i) \times \mu_{E^n \cup F^n}(f^n_j).$$

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:*

[40] Cf. (Hofer-Szabó et al., 1999).

[41] The recent experiment reported by (Weihs et al., 1998) gives strong support to this reading.

[42] 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.

1. K-multiplying results r^{ein} and r^{fjn} to obtain r_k^{ein} and r_k^{fjn}, respectively, for all $k \leq K$, $i \leq I$, $j \leq J$, and $n \leq N$;
2. extending outcomes e_i^n and f_j^n into e_i^{n*} and f_j^{n*}, respectively, for all $i \leq I, j \leq J$, and $n \leq N$;
3. requiring that for each atomic outcome $e_i^n \cap f_j^n$ of $E^n \cup F^n$
$$(r_k^{ein})^{\perp_{E^n} \perp_{E^n}} \cap (r_l^{fjn})^{\perp_{F^n} \perp_{F^n}} \neq \emptyset \text{ iff } k = l \wedge k \leq K.$$
4. adding an event C with K atomic outcomes.

The common common cause C should satisfy:

1. For all $i \leq I,\ j \leq J,\ k \leq K,\ n \leq N$,
$$C \prec r_k^{ein} \text{ and } C \prec r_k^{fjn}, \qquad \text{(causal precedence)}$$
where r_k^{ein} and r_k^{fjn} are K-multiplied results r^{ein} and r^{fjn}, respectively;
2. Every history that contains C, contains E^n and F^n for some $n \leq N$; *(conservation)*
3. For all $i \leq I, j \leq J, k \leq K, n \leq N$,
$$\overset{*}{\mu}_{E^n \cup F^n \cup C}(e_i^{n\,*} \cap f_j^{n\,*} \mid \omega_k) = \overset{*}{\mu}_{E^n \cup F^n \cup C}(e_i^{n\,*} \mid \omega_k) \times \overset{*}{\mu}_{E^n \cup F^n \cup C}(f_j^{n\,*} \mid \omega_k),$$
where $\overset{*}{\mu}_{E^n \cup F^n \cup C}(\omega_k) = \overset{*}{\mu}_{E^n \cup F^n \cup C}(\mathbf{1}_{E^n} \cap \mathbf{1}_{F^n} \cap \omega_k)$; *(screening-off)*
4. For all $i \leq I, j \leq J, k \leq K, n \leq N$,
$$\mu_{E^n \cup F^n}(e_i^n \cap f_j^n) = \overset{*}{\mu}_{E^n \cup F^n}(e_i^{n\,*} \cap f_j^{n\,*}). \qquad \text{(adequacy)}$$

Events E^n (F^N) are assumed to be incompatible since they represent alternative measurement. Although they are alternative, there is some event at which one or another setting is chosen, and this translates into requiring that event S exists. Note that Causal Precedence and Conservation guarantee that if only (in the initial model) $e_i^n \cap f_j^n \neq \emptyset$, then (in the enlarged model) for any ω_k, $e_i^{n\,*} \cap f_j^{n\,*} \cap \omega_k \neq \emptyset$. Thus, the Locality Requirement applies:
$$\overset{*}{\mu}_{E^n \cup F^n}(e_i^{n\,*} \cap f_j^{n\,*}) = \overset{*}{\mu}_{E^n \cup F^n \cup C}(e_i^{n\,*} \cap f_j^{n\,*}).$$
This allows us to formulate the condition of Adequacy.

B Proofs from section 6.2

To see that p_{E_1,E_2,F_1,F_2} returns the 'single' probabilities μ_{E_i} and μ_{F_j} as marginals, note that, e.g.,
$$\overset{*}{\mu}_{E_1 \cup C}\ (\mathbf{1}_{E_1} \cap \mathbf{1}_C \mid \omega_i) = \frac{\overset{*}{\mu}_{E_1 \cup C}\ (\mathbf{1}_{E_1} \cap \mathbf{1}_C \cap \omega_i)}{\overset{*}{\mu}_{E_1 \cup C}\ (\mathbf{1}_{E_1} \cap \omega_i)} = 1 \qquad (25)$$
and, e.g.,
$$\begin{aligned} &\sum_{i=1}^{K} \overset{*}{\mu}_{E_2 \cup C}\ (e_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_C\ (\omega_i) = \\ &\sum_{i=1}^{K} \overset{*}{\mu}_{E_2 \cup C}\ (e_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{E_2 \cup C}\ (\mathbf{1}_{E_2} \cap \omega_i) = \\ &\overset{*}{\mu}_{E_2 \cup C}\ (e_2 \cap \mathbf{1}_C) = \overset{*}{\mu}_{E_2}\ (e_2) = \mu_{E_2}(e_2), \end{aligned} \qquad (26)$$

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 'double' joints for compatible events, let us calculate as an example:

$$p_{E_1,E_2,F_1,F_2}(e_1, \mathbf{1}_{E_2}, \mathbf{1}_{F_1}, f_2) = \\ \sum_{i=1}^{K} \overset{*}{\mu}_{E_1 \cup C}\ (e_1 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{F_2 \cup C}\ (f_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_C\ (\omega_i). \tag{27}$$

By the Locality Requirement we can identify

$$\begin{aligned} \overset{*}{\mu}_{E_1 \cup C}\ (e_1 \cap \mathbf{1}_C \mid \omega_i) &= \overset{*}{\mu}_{E_1 \cup F_2 \cup C}\ (e_1 \cap \mathbf{1}_{F_2} \cap \mathbf{1}_C \mid \omega_i), \\ \overset{*}{\mu}_{F_2 \cup C}\ (f_2 \cap \mathbf{1}_C \mid \omega_i) &= \overset{*}{\mu}_{E_1 \cup F_2 \cup C}\ (\mathbf{1}_{E_1} \cap f_2 \cap \mathbf{1}_C \mid \omega_i). \end{aligned} \tag{28}$$

Screening-off then gives us

$$\overset{*}{\mu}_{E_1 \cup C}\ (e_1 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{F_2 \cup C}\ (f_2 \cap \mathbf{1}_C \mid \omega_i) = \overset{*}{\mu}_{E_1 \cup F_2 \cup C}\ (e_1 \cap f_2 \cap \mathbf{1}_C \mid \omega_i). \tag{29}$$

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

$$\overset{*}{\mu}_C\ (\omega_i) = \overset{*}{\mu}_{E_1 \cup F_2 \cup C}\ (\omega_i) = \overset{*}{\mu}_{E_1 \cup F_2 \cup C}\ (\mathbf{1}_{E_1} \cap \mathbf{1}_{F_2} \cap \omega_i). \tag{30}$$

Accordingly, by the theorem of the total probability:

$$\sum_{i=1}^{K} \overset{*}{\mu}_{E_1 \cup F_2 \cup C}(e_1 \cap f_2 \cap \mathbf{1}_C \mid \omega_i) \cdot \overset{*}{\mu}_{E_1 \cup F_2 \cup C}\ (\mathbf{1}_{E_1} \cap \mathbf{1}_{F_2} \cap \omega_i) = \\ \overset{*}{\mu}_{E_1 \cup F_2 \cup C}\ (e_1 \cap f_2 \cap \mathbf{1}_C), \tag{31}$$

and the Locality Requirement guarantees

$$\overset{*}{\mu}_{E_1 \cup F_2 \cup C}\ (e_1 \cap f_2 \cap \mathbf{1}_C) = \overset{*}{\mu}_{E_1 \cup F_2}\ (e_1 \cap f_2), \tag{32}$$

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

$$p_{E_1,E_2,F_1,F_2}(e_1, \mathbf{1}_{E_2}, \mathbf{1}_{F_1}, f_2) = \mu_{E_1 \cup F_2}(e_1 \cap f_2). \tag{33}$$

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.

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 typoscript, Katowice.

Bugajski, S. (1978). A remark on joint probability distributions for quantum observables. *Zeitschrift für Naturforschung*, 33a:1383–1385.

De Muynck, W. M. (1986). The Bell inequalities and their irrelevance to the problem of locality in quantum mechanics. *Physics Letters*, 114:65–67.

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.

Fine, A. (1989). Do correlations need to be explained? In Cushing, J. T. and McMullin, E., editors, *Philosophical Consequences of Quantum Theory*, pages 175–194. Notre Dame, IN: Notre Dame University Press.

Gleason, A. M. (1957). Measures on the closed subspaces of a Hilbert space. *Journal of Mathematics and Mechanics*, 6:885–893.

Gudder, S. (1969). Have to find. *???*

Gudder, S. (1988). Have to find. *???*

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*, forthcoming.

Huelga, S., Ferrero, M., and Santos, E. (1995). Loophole-free test of the Bell inequality. *Physical Review A*, 51:5008ff.

Kochen, S. and Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17:59–87.

Kowalski, T. and Placek, T. (1999). Outcomes in branching space-time. *British Journal for the Philosophy of Science*, forthcoming.

Landau, L. J. (1987). On the violation of Bell's inequality in quantum theory. *Physics Letters*, A120:54–56.

Malament, D. (1996). Hidden variable theories and "quantum probability". Unpublished talk at LSE, June 18th, 1996.

Pitowsky, I. (1989). *Quantum Probability — Quantum Logic*. Berlin: Springer.

Placek, T. (1999). Stochastic outcomes in branching space-time. *British Journal for the Philosophy of Science*, submitted.

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

Suppes, P. and Zanotti, M. (1976). On the determinism of hidden variable theories with strict correlation and conditional statistical independence of observables. In Suppes, P., editor, *Logic and Probability in Quantum Mechanics*, pages 445–455. Dordrecht: D. Reidel.

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.

Van Fraassen, B. C. (1989). The charybdis of realism. In Cushing, J. T. and McMullin, E., editors, *Philosophical Consequences of Quantum Theory*, pages 97–113. Notre Dame, IN: Notre Dame University Press.

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.