

The sheaf-theoretic structure of non-locality and contextuality

Samson Abramsky¹ and Adam Brandenburger²

¹ Department of Computer Science, University of Oxford, Wolfson Building, Parks Road, Oxford OX1 3QD, UK

² Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012, USA

E-mail: samson@comlab.ox.ac.uk and adam.brandenburger@stern.nyu.edu

New Journal of Physics **13** (2011) 113036 (39pp)

Received 25 June 2011

Published 28 November 2011

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/13/11/113036

Abstract. We use the mathematical language of sheaf theory to give a unified treatment of non-locality and contextuality, in a setting that generalizes the familiar probability tables used in non-locality theory to arbitrary *measurement covers*; this includes Kochen–Specker configurations and more. We show that contextuality, and non-locality as a special case, correspond exactly to *obstructions to the existence of global sections*. We describe a linear algebraic approach to computing these obstructions, which allows a systematic treatment of arguments for non-locality and contextuality. We distinguish a proper hierarchy of strengths of no-go theorems, and show that three leading examples—due to Bell, Hardy and Greenberger, Horne and Zeilinger, respectively—occupy successively higher levels of this hierarchy. A general correspondence is shown between the existence of local hidden-variable realizations using negative probabilities, and no-signalling; this is based on a result showing that the linear subspaces generated by the non-contextual and no-signalling models, over an arbitrary measurement cover, coincide. Maximal non-locality is generalized to maximal contextuality, and characterized in purely qualitative terms, as the non-existence of global sections in the support. A general setting is developed for the Kochen–Specker-type results, as generic, model-independent proofs of maximal contextuality, and a new combinatorial condition is given, which generalizes the ‘parity proofs’ commonly found in the literature. We also show how our abstract setting can be represented in quantum mechanics. This leads to a strengthening of the usual no-signalling theorem, which shows that quantum mechanics obeys no-signalling for arbitrary families of commuting observables, not just those represented on different factors of a tensor product.

Contents

1. Introduction	2
2. The setting	5
2.1. A basic scenario	5
2.2. Events	6
2.3. Distributions	6
2.4. Measurement covers	7
2.5. Empirical models	9
2.6. Examples	9
3. Global sections	10
4. Existence of global sections	12
4.1. The incidence matrix	12
4.2. Example: Bell-type scenarios	13
4.3. Global sections as solutions of linear systems	14
4.4. Examples	14
5. Negative probabilities	16
5.1. Solving the linear system over the reals	17
5.2. Example: the Popescu–Rohrlich box	19
5.3. An explicit formula for the dimension	19
5.4. Global sections and no-signaling	21
6. Strong contextuality	21
6.1. Strong contextuality and maximal non-locality	23
7. Generic strong contextuality and Kochen–Specker theorems	24
7.1. Kochen–Specker graphs	25
8. Global sections and hidden variables	27
9. Quantum representations	29
9.1. Greenberger, Horne and Zeilinger models	32
9.2. Kochen–Specker representations	33
9.3. Bell-type scenarios and Kochen–Specker theorems	34
10. Postlude	34
10.1. Related work	35
Acknowledgments	36
Appendix	37
References	37

1. Introduction

Non-locality and contextuality are fundamental features of physical theories, which contradict the intuitions underlying classical physics. In particular, they are prominent features of quantum mechanics, and the goal of the classical no-go theorems of Bell [1], Kochen–Specker [2], etc is to show that they are *necessary features* of any theory whose experimental predictions agree with those of quantum mechanics.

Bell's insights into non-locality have been seminal to current developments in quantum information, where entanglement is viewed as a key informatic resource; there has also been considerable recent work on experimental tests for contextuality [3, 4].

In this paper, we study these notions from a novel perspective, which yields new insights and results. Our approach has the following notable features.

- The importance of Bell's theorem and related results is that they apply not just to quantum mechanics, but to *all* theories with certain structural properties. We introduce a general mathematical setting, completely independent of Hilbert space, which strengthens this feature and allows results to be proved in considerable generality.
- We study non-locality and contextuality in a unified setting. The idea that non-locality can be seen as a particular form of contextuality, and specific results obtaining Bell-type non-locality from Kochen–Specker configurations, can be found in, for example, [5–7]. An important recent contribution in this direction is [8], which studies non-contextual inequalities as a generalization of Bell inequalities.

Our approach focuses on structural aspects. It offers a general, systematic and mathematically robust setting in which non-locality and contextuality are treated in a unified fashion; our definitions and results specialize to yield standard formulations of either as special cases, but subsume both.

- We use the mathematics of *sheaf theory* to analyze the structure of non-locality and contextuality. Sheaf theory is pervasive in modern mathematics, allowing the passage from local to global [9]. Starting from a simple experimental scenario, and the kind of probabilistic models familiar from discussions of Bell's theorem, the Popescu–Rohrlich (PR) boxes [10], etc, we show that there is a very direct, compelling formalization of these notions in sheaf-theoretic terms. Moreover, on the basis of this formulation, we show that the phenomena of non-locality and contextuality can be characterized precisely in terms of *obstructions to the existence of global sections*. We give linear algebraic methods for computing these obstructions.

These ideas lead, in turn, to a number of novel insights into non-locality and contextuality.

- We are able to distinguish three *strengths* of degree of non-locality: standard *probabilistic non-locality*, exhibited by the original example of Bell; *possibilistic non-locality*, exemplified by the well-known *Hardy model* [11]; and *strong contextuality*. These three properties form a strict hierarchy; strong contextuality implies possibilistic non-locality, which implies probabilistic non-locality, but the converse implications fail. In fact, we show that the Bell model is probabilistically but not possibilistically non-local; the Hardy model is possibilistically non-local but not strongly contextual; and the *Greenberger, Horne and Zeilinger (GHZ) models* [12], for the case when the number of parties is greater than two, are strongly contextual. Thus we have a hierarchy

$$\text{Bell} < \text{Hardy} < \text{GHZ}.$$

Moreover, Lal has shown [13] that the only bipartite no-signalling devices satisfying strong contextuality are the PR boxes, thus giving a new characterization of these super-quantum devices.

- We show that strong contextuality is equivalent to a quantitative notion of *maximal contextuality*, which has been studied in the special case of Bell-type scenarios as

maximal non-locality. We use this equivalence to characterize maximal contextuality, and in particular maximal non-locality, in terms of a *Boolean satisfiability problem* naturally associated with a probabilistic model, for the case of dichotomic measurements, and more generally in terms of a *constraint satisfaction problem (CSP)*.

- We apply our linear algebraic methods of constructing global sections to the issue of giving local hidden-variable realizations using *negative probabilities* [14–17]. We show that there is an equivalence between the existence of such realizations and the *no-signalling* property.
- We give a general perspective on the Kochen–Specker-type theorems as generic (model-independent) proofs of strong contextuality. We show the general combinatorial structure of these results, and make connections to graph theory, leading to the notion of *Kochen–Specker graphs*, defined in purely graph-theoretic terms.
- We prove a general result (theorem 8.1) which shows a strict equivalence between the realization of a system by a *factorizable hidden-variable model*, and the existence of a global section which glues together a certain compatible family on a presheaf. Factorizability is a general property, which subsumes both Bell locality and a form of non-contextuality at the level of distributions as special cases. This means that the whole issue of non-locality and contextuality can be translated into a canonical mathematical form, in terms of obstructions to the existence of certain global sections. This opens up the possibility of applying the powerful methods of sheaf theory to the study of the structure of these notions.
- We show in detail how the abstract setting we use can be represented in quantum mechanics; hence our results apply to all the standard situations. One interesting point that emerges from this is that the property of *compatibility* of a family of sections on a presheaf corresponds to a form of *no-signalling* [18]. This form of no-signalling subsumes, but is more general than, the usual notion; it applies to arbitrary families of commuting observables, not just to those represented on different factors of a tensor product. We therefore prove a *generalized no-signalling theorem*, showing that quantum mechanics does satisfy this more general property.

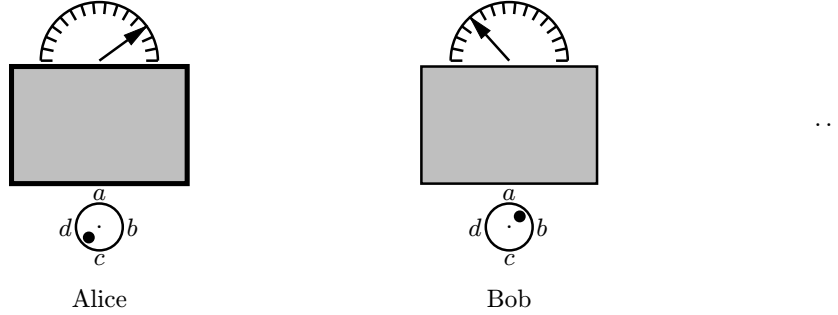
The remainder of this paper is organized as follows. The basic setting is motivated and laid out in section 2. The correspondence between global sections and (deterministic) local hidden variables is explained in section 3. The linear algebraic method for constructing global sections (or determining their non-existence) is presented in section 4, together with the results relating to the Bell and Hardy models. The equivalence between no-signalling and the existence of local hidden-variable realizations with negative probabilities is proved in section 5. Strong contextuality, the results relating to the GHZ models and the hierarchy between Bell, Hardy and GHZ, and the connections with maximal non-locality, are presented in section 6. The general combinatorial structure of Kochen–Specker-type theorems is studied in section 7. In section 8, we prove our general result relating factorizable hidden-variable models to the existence of global sections. Representations in quantum mechanics, and the generalized form of no-signaling, are treated in section 9. Section 10 contains a postlude, summarizing what has been done, discussing related work and describing some further directions.

The mathematical background needed to read this paper is quite modest. In particular, only the bare definitions of category and functor are required. A brief [appendix](#) reviews these definitions.

2. The setting

2.1. A basic scenario

Our starting point is the idealized situation depicted in the following diagram.



There are several agents or experimenters, who can each select one of several different measurements to carry out, and observe one of several different outcomes. These agents may or may not be spatially separated. When a system is prepared in a certain fashion and measurements are selected, some corresponding outcomes will be observed. These individual occurrences or ‘runs’ of the system are the basic events. Repeated runs allow relative frequencies to be tabulated, which can be summarized by a probability distribution on events for each selection of measurements. We shall call such a family of probability distributions, one for each choice of measurements, an *empirical model*.

As an example of such a model, consider the following table.

A	B	(0, 0)	(1, 0)	(0, 1)	(1, 1)
a	b	$1/2$	0	0	$1/2$
a'	b	$3/8$	$1/8$	$1/8$	$3/8$
a	b'	$3/8$	$1/8$	$1/8$	$3/8$
a'	b'	$1/8$	$3/8$	$3/8$	$1/8$

The intended scenario here is that Alice can choose between measurements a and a' , and Bob can choose b or b' . Thus the *measurement contexts* are

$$\{a, b\}, \quad \{a', b\}, \quad \{a, b'\}, \quad \{a', b'\},$$

and these index the rows of the table. Each measurement has possible outcomes 0 or 1. The matrix cell at row (a', b) and column $(0, 1)$ corresponds to the event where Alice performs a' and observes the outcome 0, and Bob performs b and observes the outcome 1. This can be described by the function

$$\{a' \mapsto 0, b \mapsto 1\}.$$

The cells of the row indexed by $\{a', b\}$ correspond to the set of functions O^C , where C is the measurement context $\{a', b\}$, and $O = \{0, 1\}$ is the set of outcomes³.

Each row of the table specifies a probability distribution on events for a given choice of measurements, i.e. on the set O^C where the row is indexed by C . For example, the event

$$\{a' \mapsto 0, b \mapsto 1\}$$

is specified to have the probability $1/8$.

³ O^C denotes the set of functions from C to O . This and a few other set-theoretic notations are explained in the [appendix](#).

The basic ingredients of our formalism will be the *measurement contexts*, the *events* and the *distributions* on events. A model of a particular measurement scenario will be given by specifying a set of measurements X , a family \mathcal{M} of measurement contexts, and for each context $C \in \mathcal{M}$, a distribution on the events O^C .

We shall now proceed to formalize these ideas. Simple as this setting may seem, it does have significant mathematical structure, which our formalization will enable us to articulate.

2.2. Events

We shall fix a set X of measurements. We shall also fix a set O of possible *outcomes* for each measurement⁴. Throughout this paper, we shall assume that X and O are finite.

For each set of measurements $U \subseteq X$, a *section over U* is a function $s : U \rightarrow O$. Such a section describes the event in which the measurements in U were carried out, and the outcome $s(m)$ was observed for each $m \in U$.

We shall write $\mathcal{E} : U \mapsto O^U$ for the assignment of the set of sections over U to each set of measurements U . There is also a natural action by restriction. If $U \subseteq U'$, there is a map

$$\text{res}_{U'}^U : \mathcal{E}(U') \rightarrow \mathcal{E}(U) :: s \mapsto s|_U.$$

Note that $\text{res}_U^U = \text{id}_U$, and if $U \subseteq U' \subseteq U''$, then

$$\text{res}_U^{U''} \circ \text{res}_{U'}^{U''} = \text{res}_U^{U'}.$$

Altogether, this says that \mathcal{E} is a *presheaf*, i.e. a functor $\mathcal{E} : \mathcal{P}(X)^{\text{op}} \rightarrow \mathbf{Set}$.

\mathcal{E} has an important additional property. Suppose we are given a family of sets $\{U_i\}_{i \in I}$ with $\bigcup_{i \in I} U_i = U$; that is, the family $\{U_i\}$ is a *cover* of U . Suppose, moreover, that we are given a family of sections $\{s_i \in \mathcal{E}(U_i)\}_{i \in I}$, which is *compatible* in the following sense: for all $i, j \in I$,

$$s_i|_{U_i \cap U_j} = s_j|_{U_i \cap U_j}.$$

Then there is a unique section $s \in \mathcal{E}(U)$ such that $s|_{U_i} = s_i$ for all $i \in I$. This says that we can glue together local data that are compatible in the sense of agreeing on overlaps; moreover, this glued section is uniquely determined.

This gluing property is known as the *sheaf condition*; it says that \mathcal{E} is a sheaf, which we shall refer to as the *sheaf of events*.

The fact that this sheaf condition holds for \mathcal{E} is quite trivial, since we are simply looking at functions on a discrete space; we can always glue together partial functions which agree on their overlaps, just by taking the union of their graphs.

2.3. Distributions

To capture the idea that empirically we observe statistical rather than deterministic behavior in microphysical systems, we shall consider distributions on events. It will be advantageous to allow some generality in the notion of distribution we shall consider, by taking the algebra of probabilistic ‘weights’ as a parameter.

A *commutative semiring* is a structure $(R, +, 0, \cdot, 1)$, where $(R, +, 0)$ and $(R, \cdot, 1)$ are commutative monoids, and moreover multiplication distributes over addition:

$$x \cdot (y + z) = x \cdot y + x \cdot z.$$

⁴ We could allow a different set of outcomes for each individual measurement, but we will not need this extra generality.

There are three main examples of semirings which will be of interest: the reals

$$(\mathbb{R}, +, 0, \times, 1),$$

the non-negative reals

$$(\mathbb{R}_{\geq 0}, +, 0, \times, 1)$$

and the Booleans

$$\mathbb{B} = (\{0, 1\}, \vee, 0, \wedge, 1).$$

We fix a semiring R . Given a set X , the *support* of a function $\phi : X \rightarrow R$ is the set of $x \in X$ such that $\phi(x) \neq 0$. We write $\text{supp}(\phi)$ for the support of ϕ . An R -distribution on X is a function $d : X \rightarrow R$ which has finite support, and is such that

$$\sum_{x \in X} d(x) = 1.$$

Note that the finite support condition ensures that this sum is well defined. We write $\mathcal{D}_R(X)$ for the set of R -distributions on X .

In the case of the semiring $\mathbb{R}_{\geq 0}$, this is the set of probability distributions with finite support on X . In the case of the Booleans \mathbb{B} , it is the set of non-empty finite subsets of X ; thus possibilistic or relational models [19, 20] are also covered. In the case of the reals \mathbb{R} , it is the set of *signed measures* with finite support, allowing for ‘negative probabilities’ [14–17].

Given a function $f : X \rightarrow Y$, we define

$$\mathcal{D}_R(f) : \mathcal{D}_R(X) \rightarrow \mathcal{D}_R(Y) :: d \mapsto \left[y \mapsto \sum_{f(x)=y} d(x) \right].$$

This is easily seen to be functorial:

$$\mathcal{D}_R(g \circ f) = \mathcal{D}_R(g) \circ \mathcal{D}_R(f), \quad \mathcal{D}_R(\text{id}_X) = \text{id}_{\mathcal{D}_R(X)},$$

so we have a functor $\mathcal{D}_R : \mathbf{Set} \rightarrow \mathbf{Set}$.⁵

We can compose this functor with the event sheaf $\mathcal{E} : \mathcal{P}(X)^{\text{op}} \rightarrow \mathbf{Set}$, to form a presheaf $\mathcal{D}_R \mathcal{E} : \mathcal{P}(X)^{\text{op}} \rightarrow \mathbf{Set}$, which assigns to each set of measurements U the set $\mathcal{D}_R(\mathcal{E}(U))$ of distributions on U -sections. It is worth writing out the functorial action of this presheaf explicitly. Given $U \subseteq U'$ we have a map

$$\mathcal{D}_R \mathcal{E}(U') \rightarrow \mathcal{D}_R \mathcal{E}(U) :: d \mapsto d|U,$$

where for each $s \in \mathcal{E}(U)$:

$$d|U(s) := \sum_{s' \in \mathcal{E}(U'), s'|U=s} d(s').$$

Thus $d|U$ is the *marginal* of the distribution d , which assigns to each section s in the smaller context U the sum of the weights of all sections s' in the larger context which restrict to s .

2.4. Measurement covers

A crucial point is that it may *not* be possible, in general, to carry out all measurements together. This is implicit in the idea that each agent makes a choice of measurement from

⁵ This functor forms part of the well-known *distribution monad*; see, e.g., [21] for references.

several alternatives; only the measurements that are chosen are actually carried out. In the situation where the agents are spatially separated, and the measurements that each carries out are localized to their own site, the measurements at the different parts can be carried out jointly. In general, we must allow for more complex situations, where compatible sets of measurements may overlap in complicated ways.

We shall now introduce the notion of measurement cover, which formalizes the idea that only certain measurements can be carried out jointly.

A *measurement cover* \mathcal{M} on the set X of measurements is a family of subsets of X such that

- $\bigcup \mathcal{M} = X$;
- \mathcal{M} is an *anti-chain*, i.e. $C, C' \in \mathcal{M}$ and $C \subseteq C'$ implies $C = C'$.

We think of X as a set of labels for the basic measurements in an experiment. A set $C \in \mathcal{M}$ is a *measurement context*, a set of measurements that can be carried out jointly. We shall focus on the *maximal* compatible sets of measurements, hence the anti-chain condition. Any compatible family of measurements in X will be included in some element of \mathcal{M} .

It should be noted that measurement covers provide a very general way of representing compatibility relationships. Of course, a physical interpretation in particular circumstances will give rise to specific structures of this kind. We shall discuss quantum representations of the formalism in detail in section 9. We also discuss the conceptual consequences of our (and related) results concerning compatibility in the postlude in section 10.

2.4.1. Bell-type scenarios. We shall now describe a particular class of measurement covers that arises in the formulation of Bell-type theorems on non-locality and in the study of PR boxes and other non-local devices.

Consider a disjoint family $\{X_i\}_{i \in I}$. We think of I as labeling the parts of a system, which may be space-like separated; X_i is the set of basic measurements which can be carried out at part i . We form the disjoint union X of this family. We define \mathcal{M} to be those subsets of X containing exactly one measurement from each part. Thus we regard measurements carried out in different parts of the system as compatible, but do not allow for compatible measurements in the same part.

2.4.2. Kochen–Specker-type scenarios. Measurement covers are general enough to cover the situations arising in Kochen–Specker style proofs of contextuality, as well as the Bell-type scenarios for non-locality.

Consider the set $X = \{m_1, \dots, m_{18}\}$, and the measurement cover \mathcal{M} whose elements are the columns of the following table:

m_1	m_1	m_8	m_8	m_2	m_9	m_{16}	m_{16}	m_{17}
m_2	m_5	m_9	m_{11}	m_5	m_{11}	m_{17}	m_{18}	m_{18}
m_3	m_6	m_3	m_7	m_{13}	m_{14}	m_4	m_6	m_{13}
m_4	m_7	m_{10}	m_{12}	m_{14}	m_{15}	m_{10}	m_{12}	m_{15}

The importance of this example is that it can be realized by unit vectors in \mathbb{R}^4 , such that each measurement context C in \mathcal{M} is an orthogonal set of vectors. This structure is used in the 18-vector proof of the Kochen–Specker theorem in [22].

We shall discuss proofs of the Kochen–Specker theorem in detail later in this paper: from the point of view of the abstract, ‘logical’ structure in section 7 and as regards the interpretation in quantum mechanics in section 9.2.

2.5. Empirical models

We shall now show how the intuitive scenario described at the beginning of this section can be captured formally, using the mathematical structure we have developed.

Suppose we are given a measurement cover \mathcal{M} . Recall that \mathcal{M} covers X , i.e. $\bigcup \mathcal{M} = X$.

We shall define a *no-signaling empirical model* for \mathcal{M} to be a compatible family for the cover \mathcal{M} with respect to the presheaf $\mathcal{D}_R\mathcal{E}$. This means that for each measurement context $C \in \mathcal{M}$, there is a distribution $e_C \in \mathcal{D}_R\mathcal{E}(C)$. Moreover, this family of distributions is compatible in the sense of the sheaf condition: for all $C, C' \in \mathcal{M}$,

$$e_C|_{C \cap C'} = e_{C'}|_{C \cap C'}.$$

In the case of Bell-type scenarios, this is readily seen to coincide with the usual notion of no-signaling. For example, in the bipartite case, consider contexts $C = \{m_a, m_b\}$, $C' = \{m_a, m'_b\}$, with a choice of measurement each for Alice and Bob. Fix $s_0 \in \mathcal{E}(\{m_a\})$, which assigns some outcome to m_a . Then the compatibility condition implies that

$$\sum_{s \in \mathcal{E}(C), s|_{m_a=s_0}} e_C(s) = \sum_{s' \in \mathcal{E}(C'), s'|_{m_a=s_0}} e_{C'}(s').$$

This says that the probability for Alice to get the outcome specified by s_0 for her measurement m_a is the same, whether we marginalize over the possible outcomes for Bob when he makes the measurement m_b , or the measurement m'_b . In other words, Bob’s choice of measurement cannot influence Alice’s outcome. This is exactly the standard definition of no-signaling.

We should also note, as a boundary case, that $\mathcal{E}(\emptyset)$ is a one-element set, and $\mathcal{D}_R\mathcal{E}(\emptyset)$ is again a one-element set. Thus if contexts have empty intersection, the compatibility condition is trivially satisfied.

The general notion of a compatible family for arbitrary covers \mathcal{M} applies to a much wider range of situations than Bell-type scenarios; later we will show that the empirical models which can be represented as quantum mechanical systems satisfy this more general form of no-signaling.

As we shall only consider no-signaling models in this paper, henceforth we shall simply speak of *empirical models*.

2.6. Examples

We shall now show how some standard examples appear in our formalism.

Consider a bipartite Bell-type scenario, where Alice has two possible measurements $\{a, a'\}$, and Bob has $\{b, b'\}$. There are two possible outcomes, 0 or 1, for each measurement.

Thus there are four maximal measurement contexts:

$$\{a, b\}, \{a', b\}, \{a, b'\}, \{a', b'\},$$

which index the rows of the following table:

	(0, 0)	(1, 0)	(0, 1)	(1, 1)
(a, b)	p_1	p_2	p_3	p_4
(a', b)	p_5	p_6	p_7	p_8
(a, b')	p_9	p_{10}	p_{11}	p_{12}
(a', b')	p_{13}	p_{14}	p_{15}	p_{16}

The rows of this table correspond to the sets of sections $\mathcal{E}(C)$, where C ranges over the maximal measurement contexts. Thus, for example, the cell labeled with p_2 corresponds to the section $\{a \mapsto 1, b \mapsto 0\}$ in $\mathcal{E}(C)$, where $C = \{a, b\}$.

The table specifies a weight p_i for each of these sections; in the standard case of probabilistic models, these will be non-negative reals, such that the values along each row sum to 1, and hence form a probability distribution. The distributions e_C for each maximal context C collectively specify what we are calling an empirical model; and the no-signalling condition corresponds exactly to the compatibility condition on this family of distributions.

As a specific example, consider the following table:

A	B	(0, 0)	(1, 0)	(0, 1)	(1, 1)
a	b	1/2	0	0	1/2
a'	b	3/8	1/8	1/8	3/8
a	b'	3/8	1/8	1/8	3/8
a'	b'	1/8	3/8	3/8	1/8

We shall use this model later to give a proof of Bell's theorem [1].

As another example, consider:

	(0, 0)	(1, 0)	(0, 1)	(1, 1)
(a, b)	1/2	0	0	1/2
(a', b)	1/2	0	0	1/2
(a, b')	1/2	0	0	1/2
(a', b')	0	1/2	1/2	0

This is a PR box [10].

We can also consider models over other semirings of weights. For example, the following is a specification of the possibilistic version of a non-local Hardy model [11], with weights in the Boolean semiring. It can be viewed as specifying the *support* of a standard probabilistic Hardy model.

	(0, 0)	(1, 0)	(0, 1)	(1, 1)
(a, b)	1	1	1	1
(a', b)	0	1	1	1
(a, b')	0	1	1	1
(a', b')	1	1	1	0

3. Global sections

We shall now show how the structures we have exposed in our mathematical description of empirical models can be brought to bear on the analysis of non-locality and contextuality.

We have already observed that the presheaf of events \mathcal{E} is in fact a sheaf; it is natural to ask if the same holds for the presheaf $\mathcal{D}_R\mathcal{E}$. Indeed, since empirical models are compatible families for this presheaf, to say that the sheaf condition holds for such a family $\{e_C\}_{C \in \mathcal{M}}$, with respect to a measurement cover \mathcal{M} , is to say that there exists a *global section* $d \in \mathcal{D}_R\mathcal{E}(X)$, defined on the entire set of measurements X . Such a global section defines a distribution on the set $\mathcal{E}(X) = O^X$, which specifies assignment of outcomes to all measurements. Moreover, this distribution must restrict to yield the probabilities specified by the empirical model on all the measurement contexts in \mathcal{M} : i.e. for all $C \in \mathcal{M}$, $d|C = e_C$. Thus the existence of a global section for the empirical model corresponds exactly to the existence of a distribution defined on all measurements, which marginalizes to yield the empirically observed probabilities. This places the idea of *extendability* of probability distributions, as studied in pioneering work by Fine [23], in a canonical and general mathematical form.

We can say more than this. A global assignment $s \in \mathcal{E}(X) = O^X$, i.e. a global section of the sheaf \mathcal{E} , can be seen as a canonical form of *deterministic hidden variable*, which assigns a definite outcome to each measurement, independent of the measurement context in which it appears. This yields an assignment $s|C$ for each measurement context C . A global section $d \in \mathcal{D}_R\mathcal{E}(X)$ specifies a distribution on this canonical set of deterministic hidden variables. Each $s \in O^X$ induces a distribution $\delta_s \in \mathcal{D}_R\mathcal{E}(X)$, where $\delta_s(s) = 1$, and $\delta_s(s') = 0$ if $s \neq s'$. The distribution induced by s on each measurement context C is $\delta_s|C$; note that $\delta_s|C = \delta_{s|C}$. Now we have

$$e_C(s) = d|C(s) = \sum_{s' \in \mathcal{E}(X), s'|C=s} d(s') = \sum_{s' \in \mathcal{E}(X)} \delta_{s'|C}(s) \cdot d(s') = \sum_{s' \in \mathcal{E}(X)} \delta_{s'|C}(s) \cdot d(s').$$

Thus the condition that $d|C = e_C$ for each measurement context C says exactly that we reproduce the empirically observed probabilities $e_C(s)$ by *averaging* over the hidden variables with respect to the distribution d .

It is also easily verified that for each context C , and $s' \in \mathcal{E}(C)$:

$$\delta_{s'|C}(s') = \prod_{x \in C} \delta_{s'|x}(s'|x).$$

That is, the probability distribution determined by s factorizes as a product of the probabilities assigned to the individual measurements, independent of the context in which they appear. We shall define a general version of this *factorizability* property later, in section 8.

If we specialize to the case of Bell-type scenarios, we see that factorizability corresponds to *Bell locality* [1]. For example, in a context $\{a, b\}$, where a is a measurement for Alice and b a measurement for Bob, then for a joint assignment of outcomes $\{a \mapsto o_1, b \mapsto o_2\}$, it says that the probability of this joint outcome determined by the hidden variable s is the product of the probabilities it determines for $\{a \mapsto o_1\}$ and $\{b \mapsto o_2\}$. In other situations, it corresponds to a form of *non-contextuality* at the level of distributions.

We can summarize this discussion as follows.

Proposition 3.1. *The existence of a global section for an empirical model implies the existence of a local (or non-contextual) deterministic hidden-variable model which realizes it⁶.*

⁶ Note that ‘deterministic hidden variable model’ means that the model is deterministic for each fixed value of the hidden variable.

We note also that, as we shall show later (see theorem 8.1), apparently more general forms of realization of empirical models by factorizable hidden-variable models, in which the hidden variables are not required to be deterministic, are in fact equivalent to canonical realizations by global sections. Thus the entire issue of non-locality and contextuality—i.e. the existence of empirical models that have no such realizations—is equivalently formulated as the non-existence of global sections for the corresponding compatible families.

Thus, we have a characterization of the phenomena of locality and non-contextuality in terms of *obstructions to the existence of global sections*, a central issue in the pervasive applications of sheaves in geometry, topology, analysis and number theory. This opens the door to the use of the powerful methods of sheaf theory in the study of non-locality and contextuality.

4. Existence of global sections

The discussion in the previous section motivates the following problem:

Given an empirical model, determine if it has a global section.

We shall give a general linear-algebraic method for answering this question, which as we have seen is equivalent to the question of whether there exists a realization of the model by local or non-contextual hidden variables.

4.1. The incidence matrix

We are given a measurement cover \mathcal{M} . The first (and main) step is to construct a matrix \mathbf{M} of 0s and 1s, which we shall call the *incidence matrix* of \mathcal{M} . This matrix is defined using only \mathcal{M} and the event sheaf \mathcal{E} , and can be applied to any empirical model expressed as a compatible family for \mathcal{M} , with respect to any distribution functor \mathcal{D}_R .

To define the incidence matrix, we firstly form the disjoint union $\coprod_{C \in \mathcal{M}} \mathcal{E}(C)$ of all the sections over the contexts in \mathcal{M} and then specify an enumeration s_1, \dots, s_p of this set. We also specify an enumeration t_1, \dots, t_q of all the global assignments $t_j \in O^X$, i.e. the global sections of the sheaf \mathcal{E} . We then form the $(p \times q)$ -matrix \mathbf{M} , with entries defined as follows:

$$\mathbf{M}[i, j] = \begin{cases} 1, & t_j|C = s_i \ (s_i \in \mathcal{E}(C)), \\ 0, & \text{otherwise.} \end{cases}$$

Conceptually, this matrix represents the tuple of restriction maps

$$\mathcal{E}(X) \longrightarrow \prod_{C \in \mathcal{M}} \mathcal{E}(C) :: s \mapsto (s|C)_{C \in \mathcal{M}}.$$

To see this, note that for each C we have the embedding

$$\mathcal{E}(C) \longrightarrow \mathcal{P}(\mathcal{E}(C)) :: s \mapsto \{s\}.$$

Thus we obtain a map $\mathcal{E}(X) \longrightarrow \prod_{C \in \mathcal{M}} \mathcal{P}(\mathcal{E}(C))$. Now we use the isomorphism

$$\prod_{i \in I} \mathcal{P}(X_i) \cong \mathcal{P}\left(\coprod_{i \in I} X_i\right)$$

to obtain a map $\mathcal{E}(X) \longrightarrow \mathcal{P}(\coprod_{C \in \mathcal{M}} \mathcal{E}(C))$. Such a map is the same thing as a relation

$$R \subseteq \mathcal{E}(X) \times \prod_{C \in \mathcal{M}} \mathcal{E}(C).$$

	(0, 0)	(1, 0)	(0, 1)	(1, 1)
(a, b)	s_1	s_2	s_3	s_4
(a', b)	s_5	s_6	s_7	s_8
(a, b')	s_9	s_{10}	s_{11}	s_{12}
(a', b')	s_{13}	s_{14}	s_{15}	s_{16}

Figure 1. Enumeration of sections.

The incidence matrix is the Boolean matrix representation of this relation. Viewing it as a 0/1-matrix over the semiring R , it acts by matrix multiplication on distributions in $\mathcal{D}_R\mathcal{E}(X)$, represented as row vectors:

$$d \mapsto (d|C)_{C \in \mathcal{M}}.$$

Thus the image of this map will be the set of families $\{e_C\}_{C \in \mathcal{M}}$ which arise from global sections.

4.2. Example: Bell-type scenarios

We shall illustrate this construction for Bell-type scenarios. Following standard terminology, we shall refer to a Bell-type scenario with n parts, each of which has k possible measurements, each with l possible outcomes, as of (n, k, l) type. Note that for a system of (n, k, l) type, there are k^n measurement contexts, for each of which there are l^n possible assignments of outcomes. Thus there are $(kl)^n$ sections over the contexts. The set of all measurements is of size kn , and there are l^{kn} global assignments. Thus the incidence matrix in this case will be of size $(kl)^n \times l^{kn}$. Each row of the matrix will contain $l^{(k-1)n}$ 1s.

We shall describe the (2,2,2) case explicitly. In this case, the matrix is 16×16 . We shall use the enumeration of sections over contexts given in the table in figure 1. We shall also use an evident enumeration of global sections obtained by viewing them as binary strings, where the i th bit indicates the assignment of an outcome to the i th measurement, listed as a, b, a', b' .

The incidence matrix is then as follows.

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

This matrix has rank 9. We shall give a general formula for the rank of the incidence matrix in proposition 5.6, and apply it to the $(n, 2, 2)$ cases in section 5.3.

We note that in the case of Bell-type scenarios, incidence matrices have been studied as ‘transfer matrices’ in [24]; see also [25].⁷ Our account generalizes this to arbitrary measurement covers and also provides a clear conceptual derivation of these matrices in terms of the restriction maps.

4.3. Global sections as solutions of linear systems

Now we consider an empirical model $\{e_C\}$, defined with respect to the distribution functor \mathcal{D}_R . Such a model assigns a weight in the semiring R to each section $s_i \in \mathcal{E}(C)$. Thus it can be specified by a vector \mathbf{V} of length p , where $\mathbf{V}[i] = e_C(s_i)$. We can also introduce a vector \mathbf{X} of length q of ‘unknowns’, one for each global section $t_j \in \mathcal{E}(X)$. Now a solution for the linear system $\mathbf{M}\mathbf{X} = \mathbf{V}$ will be a vector of values in R , one for each t_j . To ensure that this vector is a distribution, we augment \mathbf{M} with an extra row, every entry in which is 1, and similarly augment \mathbf{V} with an extra element, also set to 1. A solution for this augmented system will enforce the constraint

$$\mathbf{X}[1] + \cdots + \mathbf{X}[q] = 1$$

and hence ensure that the assignment of weights defines a distribution on $\mathcal{E}(X)$. The remaining equations ensure that this distribution restricts to yield the weight specified by the empirical model for each section s_j .

Proposition 4.1. *Let \mathbf{M}' be the augmented incidence matrix, and \mathbf{V}' the augmented vector corresponding to an empirical model e over the distribution functor \mathcal{D}_R . Solutions to this system of equations $\mathbf{M}'\mathbf{X} = \mathbf{V}'$ in R correspond bijectively to global sections for e .*

We also note that in the case of Bell-type scenarios of (n, k, l) type, it is not necessary to use the augmented system; the solutions of the equation $\mathbf{M}\mathbf{X} = \mathbf{V}$ will automatically be distributions. This follows easily from the regular structure of the incidence matrices for these cases.

4.4. Examples

We shall consider a number of examples, based on the models of $(2, 2, 2)$ type discussed in section 2.6.

4.4.1. The Bell model. We look again at the Bell model

	(0, 0)	(1, 0)	(0, 1)	(1, 1)
(a, b)	1/2	0	0	1/2
(a', b)	3/8	1/8	1/8	3/8
(a, b')	3/8	1/8	1/8	3/8
(a', b')	1/8	3/8	3/8	1/8

We are interested in finding a solution in the non-negative reals, i.e. a probability distribution on the global assignments $\mathcal{E}(X)$. This amounts to solving the linear system over

⁷ We thank one of the journal referees for bringing this connection to our attention.

the reals, subject to the constraint $\mathbf{X} \geq \mathbf{0}$, i.e. to a linear programming problem. It is easy in this case to give a direct argument that there is no such solution, and hence that the above model has no hidden-variable realization, thus proving Bell's theorem [1].

Proposition 4.2. *The Bell model has no global section*

Proof. We focus on four of the 16 equations, corresponding to rows 1, 6, 11 and 13 of the incidence matrix. We write X_i rather than $\mathbf{X}[i]$.

$$X_1 + X_2 + X_3 + X_4 = 1/2$$

$$X_2 + X_4 + X_6 + X_8 = 1/8$$

$$X_3 + X_4 + X_{11} + X_{12} = 1/8$$

$$X_1 + X_5 + X_9 + X_{13} = 1/8$$

Adding the last three equations yields

$$X_1 + X_2 + X_3 + 2X_4 + X_5 + X_6 + X_8 + X_9 + X_{11} + X_{12} + X_{13} = 3/8.$$

Since all these terms must be non-negative, the left-hand side of this equation must be greater than or equal to the left-hand side of the first equation, yielding the required contradiction. \square

4.4.2. The Hardy model. Now we consider the possibilistic version of the Hardy model, specified by the following table.

	(0, 0)	(1, 0)	(0, 1)	(1, 1)
(a, b)	1	1	1	1
(a', b)	0	1	1	1
(a, b')	0	1	1	1
(a', b')	1	1	1	0

This is obtained from a standard probabilistic Hardy model by replacing all positive entries by 1; thus it can be interpreted as the *support* of the probabilistic model.

In this case, we are interested in the existence of a solution over the Boolean semiring. This corresponds to a Boolean satisfiability problem. For example, the equation specified by the first row of the incidence matrix corresponds to the clause

$$X_1 \vee X_2 \vee X_3 \vee X_4,$$

while the fifth yields the formula

$$\neg X_1 \wedge \neg X_3 \wedge \neg X_5 \wedge \neg X_7.$$

A solution is an assignment of Boolean values to the variables which simultaneously satisfies all these formulae. Again, it is easy to see by a direct argument that no such assignment exists.

Proposition 4.3. *The possibilistic Hardy model has no global section over the Booleans.*

Proof. We focus on the four formulae corresponding to rows 1, 5, 9 and 16 of the incidence matrix:

$$\begin{aligned} &X_1 \vee X_2 \vee X_3 \vee X_4 \\ &\neg X_1 \wedge \neg X_3 \wedge \neg X_5 \wedge \neg X_7 \\ &\neg X_1 \wedge \neg X_2 \wedge \neg X_9 \wedge \neg X_{10} \\ &\neg X_4 \wedge \neg X_8 \wedge \neg X_{12} \wedge \neg X_{16} \end{aligned}$$

Since every disjunct in the first formula appears as a negated conjunct in one of the other three formulae, there is no satisfying assignment. \square

To understand the significance of this result, we note the following general fact.

Proposition 4.4. *Let \mathbf{M} be the incidence matrix for a cover \mathcal{M} , and let \mathbf{V} be the vector of non-negative reals corresponding to a probabilistic model over \mathcal{M} . Let \mathbf{V}_b be the Boolean vector obtained by replacing each nonzero element of \mathbf{V} by 1. If the system $\mathbf{MX} = \mathbf{V}$ has a solution over the non-negative reals, then the system $\mathbf{MX} = \mathbf{V}_b$ has a solution over the Booleans.*

Proof. This follows simply from the fact that the map from the non-negative reals to the Booleans which takes all nonzero elements to 1 is a semiring homomorphism. \square

It follows that, if the support of a probabilistic model has no global section with respect to the Boolean distribution functor $\mathcal{D}_{\mathbb{B}}$, then the probabilistic model itself has no global section with respect to the probability distribution functor $\mathcal{D}_{\mathbb{R}_{\geq 0}}$. Thus the argument given above implies that the probabilistic Hardy model also has no global section, and hence is non-local.

The converse of proposition 4.4 is false. Indeed, the Bell model, which as we have seen has no probabilistic global section, *does* have a Boolean global section for its support. This is easy to show directly, but also follows from the general results in [26], which show that Hardy models are complete for the (2,2,2)-type cases, and in particular that there must be at least three sections excluded from the support in order for non-locality to hold, while the Bell model has only two zero entries.

In this sense, we can say that the Hardy model satisfies a *stronger* non-locality property than the Bell model. In general, we say that a probabilistic model is *probabilistically non-extendable* if it has no global section over $\mathcal{D}_{\mathbb{R}_{\geq 0}}$, and *possibilistically non-extendable* if its support has no global section over $\mathcal{D}_{\mathbb{B}}$. We have seen that possibilistic non-extendability is strictly stronger than probabilistic non-extendability.

5. Negative probabilities

We shall now consider the question of extendability over real-valued distributions $\mathcal{D}_{\mathbb{R}}$, i.e. signed probability measures. Formally, this simply amounts to solving the linear system over the reals, with no additional constraints. Conceptually, this allows the introduction of *negative probabilities* in the extended model. Of course, these marginalize to yield standard non-negative probabilities in the measurement contexts stipulated by the empirical model. Thus the usual relative-frequency interpretation of the actually observed statistics is maintained.

The appearance of negative probabilities in quantum mechanics has a long history, which we shall sketch in the postlude (section 10). In this section, we shall prove that all empirical models are extendable with respect to signed probability measures. In fact, there is an *equivalence* between extendability under signed measures and no-signalling. Note that the class of no-signalling models is strictly larger than the empirical models of this type that arise in quantum mechanics. For example, it includes the superquantum Popescu–Rohrlich boxes [10].

The result therefore shows that negative probabilities, in themselves, cannot characterize quantum mechanics. This runs contrary to an impression that might be gained from the

literature. For example, Feynman writes: ‘The only difference between a probabilistic classical world and the equations of the quantum world is that somehow or other it appears as if the probabilities would have to go negative...’ [27, p 480]. In fact, the introduction of negative probabilities yields the entire no-signaling world.

5.1. Solving the linear system over the reals

Given an empirical model e over $\mathcal{D}_{\mathbb{R}}$, our aim is to find a global section. The existence of such a global section for e , which is represented by the real vector \mathbf{V} , reduces to the existence of a solution for the linear system $\mathbf{M}\mathbf{X} = \mathbf{V}$ over the reals, with no additional constraints.

Note that there is no semiring homomorphism from the *reals* to the Booleans. Indeed, if there were such a homomorphism h , we would have

$$0 = h(0) = h(1 + (-1)) = h(1) \vee h(-1) = 1 \vee h(-1) = 1.$$

A similar argument shows that there is no homomorphism from the reals to the non-negative reals. Thus there is no result analogous to proposition 4.4, and it is possible for the linear system to be solvable over the reals, while no solution exists over the non-negative reals or the Booleans.

We shall now show that such solutions exist for all no-signaling probabilistic empirical models, over arbitrary measurement covers. This substantially generalizes previous results, e.g. theorem 1 in [25].

We introduce some notation. We fix a standard set of outcomes $O := \{1, \dots, l\}$. Given a cover \mathcal{M} , we define the set of *partial contexts*:

$$\mathcal{U} := \{U \subseteq X \mid \exists C \in \mathcal{M}. U \subseteq C\}.$$

For each $U \in \mathcal{U}$ and $p \geq 0$, we define

$$\mathcal{E}^{(p)}(U) := \{s \in \mathcal{E}(U) \mid |s^{-1}(\{1\})| \leq p\},$$

the set of sections that map at most p measurements to the outcome 1.

Given a section s , we write $s_{m:=j}$ for the section defined by

$$s_{m:=j}(m) = j, \quad s_{m:=j}(m') = s(m'), \quad (m' \neq m).$$

Finally, given an empirical model e , we define

$$e^{(p)} := \{e_U(s) \mid U \in \mathcal{U}, s \in \mathcal{E}^{(p)}(U)\}.$$

Proposition 5.1. *Let e be a probabilistic model over \mathcal{M} . Then e is linearly determined by $e^{(0)}$.*

Proof. We shall prove that we can infer $e^{(p+1)}$ from $e^{(p)}$; the fact that e is determined by $e^{(0)}$ then follows by induction.

Consider $U \in \mathcal{U}$, $s \in \mathcal{E}^{(p)}$, and $m \in U$. Let $U' := U \setminus \{m\}$. Using compatibility,

$$e_{U'}(s|U') = \sum_j e_U(s_{m:=j}).$$

Hence

$$e_U(s_{m:=1}) = e_{U'}(s|U') - \sum_{j \neq 1} e_U(s_{m:=j}). \quad (1)$$

All the terms on the rhs of this equation are in $e^{(p)}$, while every element of $e^{(p+1)}$ can be written in the form of the lhs.

Unwinding the induction, every number $e_C(s)$ is given by a linear combination of values in $e^{(0)}$, obtained by back-substitution in (1). \square

We can consider probabilistic models as real vectors, as in the previous section. For a given cover \mathcal{M} , the dimension of the ambient vector space will be $t := \sum_{C \in \mathcal{M}} l^{|C|}$.

Proposition 5.2. *The dimension of the subspace of \mathbb{R}^t spanned by the vectors arising from probabilistic empirical models is bounded above by $D := \sum_{U \in \mathcal{U}} (l-1)^{|U|}$.*

Proof. From proposition 5.1, we know that any probabilistic empirical model is determined by a vector of D numbers. Moreover, the corresponding map $L : \mathbb{R}^D \rightarrow \mathbb{R}^t$ defined by equations (1) is linear. Hence, the subspace spanned by the probabilistic empirical models is contained in the image of L , and has dimension $\leq D$. \square

This gives us an upper bound on the dimension of the linear subspace generated by the ‘no-signalling polytope’ over the measurement cover \mathcal{M} . We shall now give a lower bound on the dimension of the linear space spanned by the non-contextual models over \mathcal{M} —i.e. those arising from global sections.

Given a cover \mathcal{M} , we have the set \mathcal{U} of partial contexts. Given $U \in \mathcal{U}$ and $s \in \mathcal{E}^{(0)}(U)$, we can define the global assignment $v_{U,s} : X \rightarrow O$:

$$v_{U,s}(m) = s(m), \quad (m \in U), \quad v_{U,s}(m) = 1, \quad (m \notin U).$$

Note that $v_{U,s} = v_{U',s'}$ implies that $U = U'$ and $s = s'$. Each such assignment $v_{U,s}$ defines a column vector $\mathbf{v}_{U,s} = \mathbf{M}[-, v_{U,s}]$. There are clearly $D = \sum_{U \in \mathcal{U}} (l-1)^{|U|}$ such assignments.

Proposition 5.3. *The set of vectors $\{\mathbf{v}_{U,s}\}_{U \in \mathcal{U}, s \in \mathcal{E}^{(0)}(U)}$ is linearly independent. Thus the dimension of the linear subspace of \mathbb{R}^t spanned by the vectors arising from global sections is bounded below by D .*

Proof. Suppose that $\sum_{U,s} \mu_{U,s} \mathbf{v}_{U,s} = \mathbf{0}$. We shall show that $\mu_{U,s} = 0$ for all U, s , by complete induction on $|X \setminus U|$.

Given some U', s' , we choose a row of the incidence matrix (C, s_0) such that $U' \subseteq C$ and $v_{U',s'}|_C = s_0$, so that $s_0|_{U'} = s'$. Note that, for any U'', s'' , $\mathbf{M}[(C, s_0), v_{U'',s'']}] = 1$ if and only if $v_{U'',s''}|_C = s_0$, if and only if $U'' \cap C = U'$, and $s''|_{U'} = s'$. If $v_{U'',s''} \neq v_{U',s'}$, we must then have $U'' \supset U'$; so by induction hypothesis, $\mu_{U'',s''} = 0$. Using the (C, s_0) component of the vector equation $\sum_{U,s} \mu_{U,s} \mathbf{v}_{U,s} = \mathbf{0}$, we conclude that $\mu_{U',s'} = 0$. \square

We now come to our main result.

Theorem 5.4. *Let \mathcal{M} be any measurement cover. The linear subspaces generated by the non-contextual and the no-signalling models over this cover coincide, with dimension D .*

Proof. Since the non-contextual models are a subset of the no-signalling models, this follows immediately from the matching lower bound on the dimension of the local subspace from

proposition 5.3, and the upper bound on the dimension of the no-signalling space from proposition 5.2. \square

As an immediate consequence of this result, we have the following.

Theorem 5.5. *Let \mathcal{M} be any measurement cover, and e a probabilistic model over this cover, with the corresponding vector $\mathbf{V} \in \mathbb{R}^t$. Then the linear system $\mathbf{M}\mathbf{X} = \mathbf{V}$ has a solution over the reals.*

We can also apply this result to the incidence matrix. Let \mathbf{M} be the incidence matrix defined over a cover \mathcal{M} and set of outcomes O .

Proposition 5.6. *The rank of \mathbf{M} , as a matrix over the reals, is D .*

Proof. The incidence matrix defines a linear map from the vector space generated by the global assignments O^X into \mathbb{R}^t . By theorem 5.4, the dimension of the image of this map is D . \square

5.2. Example: the Popescu–Rohrlich box

We consider the PR box:

	(0, 0)	(1, 0)	(0, 1)	(1, 1)
(a, b)	1/2	0	0	1/2
(a', b)	1/2	0	0	1/2
(a, b')	1/2	0	0	1/2
(a', b')	0	1/2	1/2	0

A simple solution of the linear system for the PR box is the vector

$$[1/2, 0, 0, 0, -1/2, 0, 1/2, 0, -1/2, 1/2, 0, 0, 1/2, 0, 0, 0].$$

This vector can be taken as giving a local hidden-variable realization of the PR box using negative probabilities. Similar realizations can be given for the other PR boxes.

5.3. An explicit formula for the dimension

We now consider some symmetry properties of a cover \mathcal{M} , and the associated family of partial contexts \mathcal{U} . We define $\mathcal{U}^{(j)}$ to be the set of elements of \mathcal{U} of cardinality j , $0 \leq j \leq n$. We say that \mathcal{M} is *homogeneous* if the following conditions hold.

- (i) All the contexts $C \in \mathcal{M}$ have the same number n of elements.
- (ii) Every set $U \in \mathcal{U}^{(j)}$ is a subset of the same number N_j of contexts $C \in \mathcal{M}$. Note that we must always have $N_0 = p$, where $p = |\mathcal{M}|$.

We consider some examples.

- Every (n, k, l) -type Bell scenario is homogeneous, with $p = k^n$ and $N_j = k^{n-j}$.
- The measurement cover described in section 2.4.2, corresponding to the Kochen–Specker proof from [22], is homogeneous, with $p = 9$, $n = 4$, $N_1 = 2$ and $N_j = 1$ for all $2 \leq j \leq 4$.
- Many of the constructions used in the Kochen–Specker proofs are homogeneous, for example the cover corresponding to the Peres–Mermin magic square [7, 28], which consists

of the rows and columns of the table

A	B	C
D	E	F
G	H	I

In this case, $p = 6$, $n = 3$, $N_1 = 2$ and $N_2 = N_3 = 1$.

Proposition 5.7. *Let \mathcal{M} be a homogeneous cover. Then,*

$$D = \sum_{j=0}^n \frac{\binom{n}{j} p(l-1)^j}{N_j}.$$

Proof. From the definitions,

$$D = \sum_{U \in \mathcal{U}} (l-1)^{|U|} = \sum_{j=0}^n |\mathcal{U}^{(j)}| (l-1)^j.$$

Homogeneity implies that $|\mathcal{U}^{(j)}| = p \binom{n}{j} / N_j$. □

We now apply this result to our examples.

- For (n, k, l) -type scenarios, we have

$$D = \sum_{j=0}^n \frac{\binom{n}{j} k^n (l-1)^j}{k^{n-j}} = \sum_{j=0}^n \binom{n}{j} k^j (l-1)^j.$$

Applying the binomial theorem, we obtain $D = (k \cdot (l-1) + 1)^n$. This retrieves the dimension established in [25], with the minor difference that the value given there is $D - 1$. This apparent discrepancy arises because marginalization over the empty set is excluded in [25], using the fact that, by normalization, $e_{\emptyset}(\emptyset) = 1$. However, in this case equations (1) are *affine* rather than linear.

- For the 18-vector cover, taking $l = 2$, we obtain $D = 118$. This can be compared with the dimension of the ambient vector space, which is $9 \times 2^4 = 144$.
- The corresponding value for the Peres–Mermin square is $D = 34$, with ambient dimension 48.

We can also apply this formula to the rank of the incidence matrix. For example, for $(n, 2, 2)$ -scenarios, where the incidence matrix is of size $4^n \times 4^n$, the rank is 3^n .

This formula for the rank can be made visually apparent in this case, by noting that, with a suitable choice of enumeration for the rows and columns, the incidence matrices $\mathbf{M}(n)$ have a self-similar inductive structure:

$$\mathbf{M}(1) : \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \mathbf{M}(n+1) : \begin{bmatrix} \mathbf{M}(n) & \mathbf{M}(n) & 0 & 0 \\ 0 & 0 & \mathbf{M}(n) & \mathbf{M}(n) \\ \mathbf{M}(n) & 0 & \mathbf{M}(n) & 0 \\ 0 & \mathbf{M}(n) & 0 & \mathbf{M}(n) \end{bmatrix}$$

5.4. Global sections and no-signaling

No-signalling has been built into our notion of empirical model through the requirement of compatibility of the family $\{e_C\}$. Note, however, that any family, whether compatible or not, gives rise to a linear system of equations $\mathbf{MX} = \mathbf{V}$. If this system has a solution, so that the family has a global section, it is *automatically* compatible and hence satisfies no-signaling.

Proposition 5.8. *Let $d \in \mathcal{D}_R\mathcal{E}(X)$ be a global section. Then the family $\{d|C\}_{C \in \mathcal{M}}$ is compatible.*

Proof. This follows immediately from the functoriality of restriction. For any $C, C' \in \mathcal{M}$:

$$\mathcal{D}_R(\text{res}_{C \cap C'}^C) \circ \mathcal{D}_R(\text{res}_C^X)(d) = \mathcal{D}_R(\text{res}_{C \cap C'}^C \circ \text{res}_C^X)(d) = \mathcal{D}_R(\text{res}_{C \cap C'}^X)(d)$$

and thus $(d|C)|C \cap C' = d|C \cap C'$. Similarly, $(d|C')|C \cap C' = d|C \cap C'$. Hence $d|C$ and $d|C'$ agree on their overlap. \square

Combining this result with theorem 5.5, we obtain the following theorem.

Theorem 5.9. *Probability models have local hidden-variable realizations with negative probabilities if and only if they satisfy no-signalling.*

Thus we have a striking equivalence between no-signalling models and those admitting local hidden-variable realizations with negative probabilities.

6. Strong contextuality

Consider a probability model over a cover \mathcal{M} . By proposition 4.4, if the model is extendable over $\mathcal{D}_{\mathbb{R}_{\geq 0}}$, its support is extendable over the Booleans. This means that there is a Boolean distribution d on $\mathcal{E}(X)$ which restricts to $\text{supp}(e_C)$ for every context $C \in \mathcal{M}$. Such a distribution is simply a non-empty subset S of $\mathcal{E}(X)$. The condition $d|C = \text{supp}(e_C)$ means that, for all $s \in S$, $s|C \in \text{supp}(e_C)$ for every $C \in \mathcal{M}$; moreover, every section in $\text{supp}(e_C)$ is of the form $s|C$ for some s in S .

Given an empirical model e , we define the set

$$S_e := \{s \in \mathcal{E}(X) : \forall C \in \mathcal{M}. s|C \in \text{supp}(e_C)\}.$$

Thus a consequence of the extendability of e is that S_e is non-empty.

We say that the model e is *strongly contextual* if this set S_e is empty. Whereas a global section for an empirical model e completely determines its behavior, asking for some assignment $s \in \mathcal{E}(X)$ which is *consistent* with the support of e is much weaker. The negative property that *not even one such assignment exists* is correspondingly much stronger. Indeed, the Hardy model, which as we saw in the previous section is possibilistically non-extendable, is *not* strongly contextual. The global assignment

$$\{a \mapsto 1, a' \mapsto 0, b \mapsto 1, b' \mapsto 0\}$$

is in S_e for this model. The Bell model similarly fails to be strongly contextual.

The question now arises: are there models coming from quantum mechanics which are strongly contextual in this sense?

We shall now show that the well-known GHZ models [12], of type $(n, 2, 2)$ for all $n > 2$, are strongly contextual. This will establish a strict hierarchy

$$\text{Bell} < \text{Hardy} < \text{GHZ}$$

of increasing strength of obstructions to non-contextual behavior for these salient models.

The GHZ model of type $(n, 2, 2)$ can be specified as follows. We label the two measurements at each part as $X^{(i)}$ and $Y^{(i)}$, and the outcomes as 0 and 1. For each context C , every s in the support of the model satisfies the following conditions.

- If the number of Y measurements in C is a multiple of 4, the number of 1s in the outcomes specified by s is even.
- If the number of Y measurements is $4k + 2$, the number of 1s in the outcomes is odd.

We will see later how a model with these properties can be realized in quantum mechanics.

Proposition 6.1. *The GHZ models are strongly contextual, for all $n \geq 3$.*

Proof. We consider the case when $n = 4k$, $k \geq 1$. Assume for a contradiction that we have a global section $s \in S_e$ for the GHZ model e .

If we take Y measurements at every part, the number of outcomes that are under the assignment is even. Replacing any two Y s by X s changes the residue class mod 4 of the number of Y s, and hence must result in the opposite parity for the number of 1 outcomes under the assignment. Thus for any $Y^{(i)}$, $Y^{(j)}$ assigned the *same* value, if we substitute X s in those positions they must receive *different* values under s . Similarly, for any $Y^{(i)}$, $Y^{(j)}$ assigned different values, the corresponding $X^{(i)}$, $X^{(j)}$ must receive the same value.

Suppose first that not all $Y^{(i)}$ are assigned the same value by s . Then, for some i, j, k , $Y^{(i)}$ is assigned the same value as $Y^{(j)}$, and $Y^{(j)}$ is assigned a different value than $Y^{(k)}$. Thus $Y^{(i)}$ is also assigned a different value than $Y^{(k)}$. Then, $X^{(i)}$ is assigned the same value as $X^{(k)}$, and $X^{(j)}$ is assigned the same value as $X^{(k)}$. By transitivity, $X^{(i)}$ is assigned the same value as $X^{(j)}$, yielding a contradiction.

The remaining cases are where all Y s receive the same value. Then any pair of X s must receive different values. But taking any three X s, this yields a contradiction, since there are only two values, so some pair must receive the same value.

The case when $n = 4k + 2$, $k \geq 1$, is proved in the same fashion, interchanging the parities. When $n \geq 5$ is odd, we start with a context containing one X , and again proceed similarly.

The most familiar case, for $n = 3$, does not admit this argument, which relies on having at least four Y s in the initial configuration. However, for this case one can easily adapt the well-known argument of Mermin using ‘instruction sets’ [29] to prove strong contextuality. This uses a case analysis to show that there are eight possible global sections satisfying the parity constraint on the three measurement combinations with two Y s and one X ; and all of these violate the constraint for the XXX measurement. \square

We shall also mention an elegant result due to Lal [13].

Proposition 6.2 (Lal). *The only strongly contextual no-signaling models of type $(2, 2, 2)$ are the PR boxes.*

Thus strong contextuality actually *characterizes* the PR boxes.

6.1. Strong contextuality and maximal non-locality

The property of strong contextuality is defined in a simple ‘qualitative’ fashion, in terms of the support of a model. As we shall now see, for probabilistic models it is equivalent to a notion which can be defined in quantitative terms and has been studied in this form in the special case of Bell-type scenarios⁸.

Suppose that $\{e_C\}_{C \in \mathcal{M}}$ is a model over the presheaf $\mathcal{D}_{\mathbb{R}_{\geq 0}}$. We consider convex decompositions

$$e = \lambda L + (1 - \lambda)q, \quad 0 \leq \lambda \leq 1, \quad (2)$$

where L is a local model and q a no-signalling model. This means that, for every $C \in \mathcal{M}$, and $s \in \mathcal{E}(C)$, we have

$$e_C(s) = \lambda L_C(s) + (1 - \lambda)q_C(s).$$

We define the *non-contextual fraction* of e to be the supremum over all λ appearing in such convex decompositions (2). This notion was introduced for Bell-type scenarios in [30]; see also [31, 32], where the terminology *local fraction* is used. A model with local fraction 0 is defined to be *maximally non-local*. Geometrically, this corresponds to the model being on a face of the no-signaling polytope with no local elements.

In the general setting of models defined on arbitrary measurement covers, we say that a model e is *maximally contextual* if the non-contextual fraction of e is 0.

Proposition 6.3. *A model e is strongly contextual if and only if it is maximally contextual.*

Proof. Suppose firstly that e admits a convex decomposition (2). By the results of section 3 and also theorem 8.1, we can take L to be a convex sum of deterministic models $\sum_i \mu_i \delta_{s_i}$, where each $s_i \in \mathcal{E}(X)$ is a global assignment. If $\lambda > 0$, then from (2), $s_i \in S_e$ for each i . Thus strong contextuality implies maximal contextuality.

For the converse, suppose that $s \in S_e$. Taking $L = \lambda \cdot \delta_s$, we shall define q such that (2) holds. For each $C \in \mathcal{M}$ and $s' \in \mathcal{E}(C)$:

$$q_C(s') := \frac{e_C(s') - \lambda \cdot \delta_{s|C}(s')}{1 - \lambda}.$$

It is easily verified that, for each C , $\sum_{s' \in \mathcal{E}(C)} q_C(s') = 1$. To ensure that q is always non-negative, we must have $\lambda \leq \inf_{C \in \mathcal{M}} e_C(s|C)$. Since this is the infimum of a finite set of positive numbers, we can find $\lambda > 0$ satisfying this condition.

It remains to verify that q is no-signalling, i.e. that $\{q_C\}$ forms a compatible family. Given $C, C' \in \mathcal{M}$, fix $s_0 \in \mathcal{E}(C \cap C')$. Now

$$q_C|C \cap C'(s_0) = 1/(1 - \lambda) \left[\left(\sum_{s' \in \mathcal{E}(C), s'|C \cap C' = s_0} e_C(s') \right) - \lambda \cdot \delta_{s|C \cap C'}(s_0) \right].$$

A similar analysis applies to $q_{C'}|C \cap C'(s_0)$. Using the compatibility of e , we conclude that $q_C|C \cap C' = q_{C'}|C \cap C'$. \square

We can use this equivalence to give a characterization of maximal contextuality, and in particular of maximal non-locality, in terms of a CSP. In the case of dichotomic measurements, this reduces to a *Boolean satisfiability problem*.

⁸ We thank one of the journal referees for pointing out this connection.

We recall that a CSP [33, 34] is specified by a triple (V, K, \mathcal{R}) , where V is a finite set of variables, K is a finite set of values and \mathcal{R} is a finite set of constraints. A constraint is a pair (C, S) , where $C \subseteq V$ and $S \subseteq K^C$. (It is more common to define a constraint as an ordered list of k variables, and a set of k -tuples of values, but this is obviously equivalent to the version given.) An assignment $s : V \rightarrow K$ satisfies a constraint (C, S) if $s|_C \in S$. A solution of the CSP (V, K, \mathcal{R}) is an assignment $s : V \rightarrow K$ which satisfies every constraint in \mathcal{R} .

Let e be a model over a cover \mathcal{M} , with outcome set O . For each $C \in \mathcal{M}$, we have the set $\mathcal{S}_e(C) := \text{supp}(e_C) \subseteq O^C$. We can associate the CSP $(X, O, \{\mathcal{S}_e(C) \mid C \in \mathcal{M}\})$ with e .

Proposition 6.4. *A probabilistic model e is maximally contextual if and only if the corresponding CSP has no solution. In particular, for Bell-type scenarios, e is maximally non-local if and only if the corresponding CSP has no solution.*

Proof. This follows directly from proposition 6.3, since global sections in the support of e are clearly in bijective correspondence to solutions of the associated CSP. \square

In the case of dichotomic measurements, the CSP reduces to a Boolean satisfiability problem. In this case, we interpret the two possible outcomes as truth values, and X as a set of propositional variables.

Given a model e , we define the formula

$$\phi_e := \bigwedge_{C \in \mathcal{M}} \bigvee_{s \in \mathcal{S}_e(C)} \psi_s$$

where for a section $s \in O^C$ we define the corresponding formula

$$\psi_s := \bigwedge_{m \in C, s(m)=\text{true}} m \wedge \bigwedge_{m \in C, s(m)=\text{false}} \neg m.$$

Proposition 6.5. *A probabilistic model e with dichotomic measurements is maximally contextual if and only if the corresponding formula ϕ_e is unsatisfiable. In particular, for Bell-type scenarios, e is maximally non-local if and only if ϕ_e is unsatisfiable.*

7. Generic strong contextuality and Kochen–Specker theorems

Let e and e' be models, such that the support of e is included in the support of e' . Then \mathcal{S}_e is included in $\mathcal{S}_{e'}$; hence if e' is strongly contextual, so is e . Thus by showing strong contextuality for a single model, we can show it for a whole class of models.

We shall fix our set of outcomes as $\{0, 1\}$. This means that we can define subsets of $\mathcal{E}(C)$ by formulae ϕ_C , with the elements of C used as propositional variables. A section $s : C \rightarrow \{0, 1\}$ can be viewed as a Boolean assignment for these variables, and ϕ_C defines the set of its satisfying assignments.

We are interested, in particular, in the formula

$$\text{ONE}(C) := \bigvee_{m \in C} (m \wedge \bigwedge_{m' \in C \setminus \{m\}} \neg m').$$

This is satisfied by those assignments with exactly one outcome set to 1.

A Kochen–Specker-type result [2] can be factored into two parts.

- (i) Defining covers \mathcal{M} such that there is no global section $s \in \mathcal{E}(X)$ which satisfies the formula

$$\phi_{\mathcal{M}} := \bigwedge_{C \in \mathcal{M}} \text{ONE}(C).$$

- (ii) Providing quantum representations for these covers, which interpret the measurements by quantum observables in such a way that *every* quantum model for this set of observables has its support included in $\text{ONE}(C)$ for each $C \in \mathcal{M}$, and hence is strongly contextual.

We shall explain the quantum aspects in a later section. Here we shall investigate the combinatorial structures involved in the first part.

We shall give a simple combinatorial condition on the cover \mathcal{M} , which is implied by the existence of a global section s satisfying $\phi_{\mathcal{M}}$. Violation of this condition therefore suffices to prove that no such global section exists.

For each $m \in X$, we define

$$\mathcal{M}(m) := \{C \in \mathcal{M} \mid m \in C\}.$$

Proposition 7.1. *If $\phi_{\mathcal{M}}$ has a global section, then every common divisor of $\{|\mathcal{M}(m)| \mid m \in X\}$ must divide $|\mathcal{M}|$.*

Proof. Suppose that there is a global section $s : X \rightarrow \{0, 1\}$ satisfying $\phi_{\mathcal{M}}$. Consider the set $X' \subseteq X$ of those m such that $s(m) = 1$. Exactly one element of X' must occur in every $C \in \mathcal{M}$. Hence there is a partition of \mathcal{M} into the subsets $\mathcal{M}(m)$ indexed by the elements of X' . Thus

$$|\mathcal{M}| = \sum_{m \in X'} |\mathcal{M}(m)|.$$

It follows that, if there is a common divisor of the numbers $|\mathcal{M}(m)|$, it must divide $|\mathcal{M}|$. \square

For example, if every $m \in X$ appears in an even number of elements of \mathcal{M} , while \mathcal{M} has an odd number of elements, then $\phi_{\mathcal{M}}$ has no global section. This corresponds to the ‘parity proofs’ which are often used in verifying Kochen–Specker-type results [22, 35].

The simplest example of this situation is the ‘triangle’, i.e. the measurement cover with elements

$$\{a, b\}, \{b, c\}, \{a, c\}.$$

This example has also been discussed, in a somewhat different context, in [36].

An example where X has 18 elements, and there are nine maximal compatible sets, each with four elements, such that each element of X is in two of these, appears in the 18-vector proof of the Kochen–Specker theorem in [22].

7.1. Kochen–Specker graphs

The measurement covers that can be represented by quantum systems are of a particular form: they are generated by a *symmetric binary compatibility relation*, since compatibility in quantum mechanics means that the observables pairwise commute. Thus, for example, the ‘triangle’ cannot arise from quantum observables.

This suggests that we should take into account this feature. It turns out that this leads us directly to some standard notions in graph theory.

An undirected graph G is specified by a finite set of vertices V_G , and a set of edges E_G , which are two-element subsets of V_G . A *clique* of G is a set $C \subseteq V_G$ with an edge between every pair of vertices in C . The set of maximal cliques of G forms a measurement cover \mathcal{M}_G .

Let G be a graph. A set $S \subseteq V_G$ is called a *stable transversal* [37] if for every maximal clique C of G (i.e. for every $C \in \mathcal{M}_G$), $|S \cap C| = 1$. Note that it is necessarily the case that a stable transversal is *independent*, i.e. there is no edge between any pair of elements of S , since otherwise we could extend this pair to a maximal clique containing both.

Proposition 7.2. *Let G be a graph. The formula $\phi_{\mathcal{M}}$ defined on \mathcal{M}_G has a global section if and only if G has a stable transversal.*

Proof. Suppose that $\phi_{\mathcal{M}}$ has a global section s . Then $T := \{m \in X | s(m) = 1\}$ is a stable transversal of G .

Conversely, suppose that T is a stable transversal of G . If we define s as the characteristic function of T on X , then $s \models \text{ONE}(C)$ for each maximal clique C of G , and so $\phi_{\mathcal{M}}$ has a global section. \square

In order to apply graph-theoretic results to the quantum situation, we need to know which graphs can arise from families of quantum observables. For reasons that will be explained when we discuss quantum representations in section 9, we are interested in graphs that can be labelled by vectors in \mathbb{R}^d , such that two vertices are adjacent if and only if the corresponding vectors are orthogonal. It turns out that in graph theory, the complementary notion is used [38], so we shall say that such graphs have a *faithful orthogonal co-representation* in \mathbb{R}^d . We must also require that the maximal cliques all have size d .

Thus we define *Kochen–Specker graphs* to be finite graphs G such that

- G has a faithful orthogonal co-representation in \mathbb{R}^d ;
- the maximal cliques of G all have the same size d ;
- G has no stable transversal.

Any such graph generates a measurement cover \mathcal{M}_G such that the formula $\phi_{\mathcal{M}_G}$ has no global section; and every such graph can be realized by quantum observables, as will be shown in section 9. Thus, these graphs provide explicit finite witnesses for generic strong contextuality. An example is provided by the orthogonality graph for \mathbb{R}^4 defined by the set of 18 vectors given in [22], as well as the various sets of 31 or more vectors that have been found in \mathbb{R}^3 [2, 39–41].

A final desideratum is to provide a purely graph-theoretic condition for the existence of a faithful orthogonal co-representation. In [38, 42] the following result is proved.

Theorem 7.3. *Every graph on n nodes whose complementary graph is $(n - d)$ -connected has a faithful orthogonal co-representation in \mathbb{R}^d .*

We note that a general graph-theoretic approach to contextuality, on somewhat different lines than ours, has been developed in [8]. Interesting connections are shown between contextuality and Lovasz’s ϑ -function [43].

In [8], ‘compatibility structures’ are studied as sets of *events* rather than measurements. This leads to the formalism of *convex operational theories* as ‘generalized probability theories’.

By contrast, we distinguish between measurements, outcomes and events. This allows the functorial presheaf structure of probabilistic models to be articulated. Moreover, we use standard probability theory, as encapsulated in the distribution functor $\mathcal{D}_{\mathbb{R}_{\geq 0}}$. The non-classical, contextual features of models arise from their functorial variation over contexts. At the same time, this mathematical structure directly reflects the basic operational scenario described at the beginning of section 2.

8. Global sections and hidden variables

We shall now consider a general notion of hidden-variable model, and show that an empirical model is realized by a factorizable hidden-variable model if and only if it has a global section.

We are given a measurement cover \mathcal{M} . We fix a set Λ of values for a hidden variable. A hidden-variable model h over Λ assigns, for each $\lambda \in \Lambda$ and $C \in \mathcal{M}$, a distribution $h_C^\lambda \in \mathcal{D}_R \mathcal{E}(C)$. It also assigns a distribution $h_\Lambda \in \mathcal{D}_R(\Lambda)$ on the hidden variables. Note that this distribution is independent of the context; this is the standard structural assumption of λ -independence [44]. We require that for each $\lambda \in \Lambda$, the family $\{h_C^\lambda\}_{C \in \mathcal{M}}$ is compatible, i.e. for all $C, C' \in \mathcal{M}$:

$$h_C^\lambda|_{C \cap C'} = h_{C'}^\lambda|_{C \cap C'}.$$

Just as compatibility for empirical models corresponds to no-signaling, compatibility for hidden-variable models corresponds to the *parameter independence* condition [45, 46].

We say that a hidden-variable model h *realizes* an empirical model e if the probabilities specified by e are recovered by averaging over the values of the hidden variable. Formally, this says that for all $C \in \mathcal{M}$ and $s \in \mathcal{E}(C)$:

$$e_C(s) = \sum_{\lambda \in \Lambda} h_C^\lambda(s) \cdot h_\Lambda(\lambda).$$

The intended purpose of hidden-variable models is to explain the non-intuitive behavior of empirical models, in particular those arising from quantum mechanics, by showing that it can be reproduced by a model whose behavior is more intuitive, at the cost of introducing hidden variables. In particular, one would like to explain the non-local and contextual behavior predicted by quantum mechanics in this way. The general property which a hidden-variable model should satisfy in order to provide such an explanation is *factorizability*, which subsumes both Bell locality [1] and a form of non-contextuality at the level of distributions. It is defined as follows.

We say that a hidden-variable model h is *factorizable* if, for every $C \in \mathcal{M}$, and $s \in \mathcal{E}(C)$:

$$h_C^\lambda(s) = \prod_{m \in C} h_C^\lambda|_{\{m\}}(s|_{\{m\}}).$$

This says that the probability assigned to a joint outcome factors as the product of the probabilities assigned to the individual measurements. Note, in particular, that if $m \in C \cap C'$, then the compatibility condition on h implies that $h_C^\lambda|_{\{m\}} = h_{C'}^\lambda|_{\{m\}}$. Thus the probability distributions on outcomes of individual measurements are independent of the contexts in which they occur.

For Bell-type scenarios, factorizability corresponds exactly to Bell locality [1]. More generally, it asserts non-contextuality at the level of distributions.

Our main result can now be stated as follows.

Theorem 8.1. *Let e be an empirical model defined on a measurement cover \mathcal{M} for a distribution functor \mathcal{D}_R . The following are equivalent.*

- (i) *e has a realization by a factorizable hidden-variable model.*
- (ii) *e has a global section.*

Proof. Proposition 3.1 shows that (ii) implies (i). It remains to prove the converse.

Suppose that e is realized by a factorizable hidden-variable model h . For each $m \in X$, we define $h_m^\lambda := h_C^\lambda|_{\{m\}} \in \mathcal{D}_R\mathcal{E}(\{m\})$ for any $C \in \mathcal{M}$ such that $m \in C$. By the compatibility of the family $\{h_C^\lambda\}$, this definition is independent of the choice of C . Also, we shall write $s|m$ rather than $s|\{m\}$. We define a distribution $h_X^\lambda \in \mathcal{D}_R\mathcal{E}(X)$ for each $\lambda \in \Lambda$ by

$$h_X^\lambda(s) = \prod_{m \in X} h_m^\lambda(s|m).$$

We must show that this is a distribution. We enumerate the set of measurements X as $X = \{m_1, \dots, m_p\}$. A global assignment $s \in \mathcal{E}(X)$ can be specified by a tuple (o_1, \dots, o_p) , where $o_i = s(m_i)$. Now we can calculate:

$$\begin{aligned} & \sum_{s \in \mathcal{E}(X)} \prod_{m \in X} h_m^\lambda(s|m) \\ &= \sum_{o_1, \dots, o_p} \prod_{i=1}^p h_{m_i}^\lambda(s|m_i) \\ &= \sum_{o_1} h_{m_1}^\lambda(m_1 \mapsto o_1) \cdot \left(\sum_{o_2} h_{m_2}^\lambda(m_2 \mapsto o_2) \cdot \left(\cdots \left(\sum_{o_p} h_{m_p}^\lambda(m_p \mapsto o_p) \right) \cdots \right) \right) \\ &= \sum_{o_1} h_{m_1}^\lambda(m_1 \mapsto o_1) \cdot \left(\sum_{o_2} h_{m_2}^\lambda(m_2 \mapsto o_2) \cdot (\cdots (1) \cdots) \right) \\ &= \vdots \\ &= \sum_{o_1} h_{m_1}^\lambda(m_1 \mapsto o_1) \cdot 1 = 1. \end{aligned}$$

We now show that for each context C in \mathcal{M} , $h_X^\lambda|_C = h_C^\lambda$. We choose an enumeration of X such that $C = \{m_1, \dots, m_q\}$, $q \leq p$.

$$\begin{aligned} h_X^\lambda|_C(s) &= \sum_{s' \in \mathcal{E}(X), s'|_C = s} h_X^\lambda(s') \\ &= \sum_{(o_1, \dots, o_p), s=(o_1, \dots, o_q)} \prod_{i=1}^p h_{m_i}^\lambda(m_i \mapsto o_i) \\ &= \prod_{i=1}^q h_{m_i}^\lambda(m_i \mapsto s(m_i)) \cdot \left(\sum_{o_{q+1}, \dots, o_p} \prod_{j=q+1}^p h_{m_j}^\lambda(m_j \mapsto o_j) \right) \\ &= h_C^\lambda(s) \cdot 1 = h_C^\lambda(s). \end{aligned}$$

Now we define a distribution $d \in \mathcal{D}_R \mathcal{E}(X)$ by averaging over the hidden variables:

$$d(s) := \sum_{\lambda \in \Lambda} h_X^\lambda(s) \cdot h_\Lambda(\lambda).$$

We verify that this is a distribution:

$$\begin{aligned} \sum_{s \in \mathcal{E}(X)} d(s) &= \sum_{\lambda \in \Lambda} \sum_{s \in \mathcal{E}(X)} h_X^\lambda(s) \cdot h_\Lambda(\lambda) \\ &= \sum_{\lambda \in \Lambda} h_\Lambda(\lambda) \cdot \left(\sum_{s \in \mathcal{E}(X)} h_X^\lambda(s) \right) \\ &= \sum_{\lambda \in \Lambda} h_\Lambda(\lambda) \cdot 1 = 1. \end{aligned}$$

It remains to show that d restricts at each context C to yield e_C .

$$\begin{aligned} d|_C(s) &= \sum_{s' \in \mathcal{E}(X), s'|_C=s} d(s') \\ &= \sum_{s' \in \mathcal{E}(X), s'|_C=s} \sum_{\lambda \in \Lambda} h_X^\lambda(s') \cdot h_\Lambda(\lambda) \\ &= \sum_{\lambda \in \Lambda} h_\Lambda(\lambda) \cdot h_X^\lambda|_C(s) \\ &= \sum_{\lambda \in \Lambda} h_\Lambda(\lambda) \cdot h_C^\lambda(s) \\ &= e_C(s). \end{aligned}$$

Thus d is a global section for e . □

This result provides a definitive justification for equating the phenomena of non-locality and contextuality with obstructions to the existence of global sections.

9. Quantum representations

Since our aim is to investigate the *general properties* of systems and physical theories, it has been important that our entire discussion has been conducted without presupposing quantum mechanics, Hilbert spaces, etc. The mathematical structures which we *have* used have arisen in a rather transparent fashion from the basic experimental scenario with which we began.

However, it is important to make explicit how the structures we have described can be represented in quantum mechanics.

We begin with measurement covers. A quantum representation of a measurement cover on a set X can be described as follows. We fix a Hilbert space \mathcal{H} . As usual, an *observable* is a bounded self-adjoint operator A on \mathcal{H} . Two observables A, B are *compatible* if they commute: $AB = BA$. In this case, the composite AB is again self-adjoint and hence forms an observable.

Given a set $\mathcal{X} = \{A_x\}_{x \in X}$ of observables on \mathcal{H} indexed by X , we form a measurement cover by taking \mathcal{M} to be the set of all maximal commuting subsets of \mathcal{X} . Note that pairwise commutation implies that the observables in each such subset, composed in any order, form

a well-defined observable. We say that an abstract measurement cover \mathcal{M} has a *quantum representation* if it arises in this way.

For Bell-type scenarios, a quantum representation will have a particular form, reflecting the usual idea that the measurements are carried out at a number of different sites, which may be space-like separated. We will have a family $\{\mathcal{H}_i\}$ of Hilbert spaces, one for each part. The elements of X_i will index a family \mathcal{X}_i of incompatible (i.e. non-commuting) observables on \mathcal{H}_i . We change these into *local observables* on the *compound system* $\mathcal{H} = \mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_k$ by defining $A^i := I \otimes \cdots \otimes A \otimes \cdots \otimes I$ for each $A \in \mathcal{X}_i$. Then, A^i commutes with B^j whenever $i \neq j$, and we can form a measurement cover of Bell type on the compound system.

It is interesting in this connection to discuss a result due to Tsirelson [47]. This result can be stated, following [48], as follows.

Theorem 9.1. *Let $\{X_i\}$ and $\{Y_j\}$ be finite, commuting sets of positive operators on a Hilbert space \mathcal{H} , generating finite-dimensional von Neumann sub-algebras of $\mathbf{B}(\mathcal{H})$. Then there exist finite-dimensional Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 such that $\{X_i\}$ can be mapped isomorphically into the sub-algebra of operators on $\mathcal{H}_1 \otimes \mathcal{H}_2$ of the form $A \otimes I$, and $\{Y_j\}$ can be mapped isomorphically into the sub-algebra of operators of the form $I \otimes B$.*

The import of this result is that, under the stated assumptions, tensor product structure can be retrieved automatically as a special case of the general situation of commuting operators on a single space. Thus, the special form of representation for Bell-type scenarios is not really necessary, although it is the one which is commonly used.

Now we turn to events. For simplicity, we shall confine ourselves to the finite-dimensional case. Recall that a self-adjoint operator A has a *spectral decomposition*

$$A = \sum_{i \in I} \alpha_i \mathbf{P}_i,$$

where α_i is the i th eigenvalue, and \mathbf{P}_i is the projector onto the corresponding eigenspace. Measuring a quantum state ρ with this observable will result in one of the observable outcomes α_i , with probability $\text{Tr}(\rho \mathbf{P}_i)$, while the state will be projected into the corresponding eigenspace.

For simplicity of notation, we shall focus on *dichotomic quantum observables*, i.e. self-adjoint operators on a Hilbert space \mathcal{H} with a spectral resolution into two orthogonal subspaces. In this case, we can use a standard two-element set $O = \{0, 1\}$ to label these outcomes, and the sheaf \mathcal{E} to record the collective outcomes of a compatible set of observables.

Thus for each basic measurement label m in X , we have an observable A_m with spectral decomposition $A_m = \alpha_m^0 \mathbf{P}_m^0 + \alpha_m^1 \mathbf{P}_m^1$, where $\mathbf{P}_m^0 + \mathbf{P}_m^1 = I$. Given a maximal set of commuting observables $C = \{A_{m_1}, \dots, A_{m_k}\}$, for each $s \in O^C$ we have a projector $\mathbf{P}_s = \mathbf{P}_{m_1}^{s(m_1)} \cdots \mathbf{P}_{m_k}^{s(m_k)}$. The composed observable $A_C = A_{m_1} \cdots A_{m_k}$ has a decomposition of the form

$$A_C = \sum_{s \in O^C} \alpha_s \mathbf{P}_s,$$

where $\alpha_s = \prod_i \alpha_{m_i}^{s(m_i)}$. To ensure that these eigenvalues can be associated with distinct outcomes, it is necessary that $\alpha_s = \alpha_t$ implies that $s = t$. This can be achieved by appropriate choices of the eigenvalues α_m^i .

It may well be the case that this decomposition contains redundant terms, in the sense that $\mathbf{P}_s = \mathbf{0}$ for some values of s . The important point is that these projectors do yield a resolution of

the identity:

$$\sum_{s \in O^C} \mathbf{P}_s = (\mathbf{P}_{m_1}^0 + \mathbf{P}_{m_1}^1) \cdots (\mathbf{P}_{m_k}^0 + \mathbf{P}_{m_k}^1) = I^k = I.$$

Now we consider empirical models. Suppose we are given an empirical model e on a measurement cover \mathcal{M} , which has a quantum representation in the form described above, based on a Hilbert space \mathcal{H} . A quantum representation of e is given by a state ρ on \mathcal{H} . For each compatible set of observables $C \in \mathcal{M}$, the state defines a probability distribution ρ_C on $\mathcal{E}(C)$, by the standard ‘statistical algorithm’ of quantum mechanics: $\rho_C(s) = \text{Tr}(\rho \mathbf{P}_s)$. Thus $\rho_C \in \mathcal{D}_{\mathbb{R}_{\geq 0}} \mathcal{E}(C)$ for each $C \in \mathcal{M}$.

An interesting point now arises: do the distributions $\{\rho_C\}$ necessarily form a compatible family? In the case of a Bell-type scenario, the fact that they do is the content of the standard *no-signalling theorem* [17]. However, Bell-type scenarios are very special cases of measurement covers. We shall therefore verify explicitly that the distributions determined by a quantum state, with respect to any family of sets of commuting observables, do form a compatible family in the sense of sheaf theory. We can regard this as a generalized form of no-signaling theorem.

Proposition 9.2 (Generalized no-signaling). *The family of distributions $\{\rho_C\}$ on families of commuting observables defined by a quantum state ρ are compatible on overlaps: for all C, C' :*

$$\rho_C|_{C \cap C'} = \rho_{C'}|_{C \cap C'}.$$

Proof. Firstly, we define $C_0 := C \cap C'$, $C_1 := C \setminus C_0$, and $C_2 := C' \setminus C_0$. Thus C is the disjoint union of C_1 and C_0 , and C' is the disjoint union of C_2 and C_0 . Note that $\mathcal{E}(C) \cong \mathcal{E}(C_0) \times \mathcal{E}(C_1)$, and $\mathcal{E}(C') \cong \mathcal{E}(C_0) \times \mathcal{E}(C_2)$. Thus we can write $s \in \mathcal{E}(C)$ as $s = (s_0, s_1)$, and similarly for sections in $\mathcal{E}(C')$. In this notation, $\mathbf{P}_{(s_0, s_1)} = \mathbf{P}_{s_0} \mathbf{P}_{s_1}$. Now we can calculate

$$\begin{aligned} \rho_C|_{C_0}(s_0) &= \sum_{s_1 \in \mathcal{E}(C_1)} \rho_C(s_0, s_1) \\ &= \sum_{s_1 \in \mathcal{E}(C_1)} \text{Tr}(\rho \mathbf{P}_{(s_0, s_1)}) \\ &= \sum_{s_1 \in \mathcal{E}(C_1)} \text{Tr}(\rho \mathbf{P}_{s_0} \mathbf{P}_{s_1}) \\ &= \text{Tr}\left(\sum_{s_1 \in \mathcal{E}(C_1)} \rho \mathbf{P}_{s_0} \mathbf{P}_{s_1}\right) \\ &= \text{Tr}(\rho \mathbf{P}_{s_0} \sum_{s_1 \in \mathcal{E}(C_1)} \mathbf{P}_{s_1}) \\ &= \text{Tr}(\rho \mathbf{P}_{s_0} I) \\ &= \text{Tr}(\rho \mathbf{P}_{s_0}) \\ &= \rho_{C_0}(s_0). \end{aligned}$$

A similar computation shows that $\rho_{C'}|_{C_0}(s_0) = \rho_{C_0}(s_0)$. Hence $\rho_C|_{C \cap C'} = \rho_{C'}|_{C \cap C'}$. \square

Thus, we see that quantum mechanics obeys a general form of no-signalling, which applies to compatible families of observables in general, not just those represented as operating on different factors of a tensor product. This form of no-signaling says that, *at the level of distributions*, the statistics obtained for a measurement on a given state are independent of the context of other compatible measurements which may also have been carried out.

We can in fact use Tsirelson's theorem 9.1 to reduce this result to the standard form of no-signalling theorem.⁹ Given sets of commuting observables C and C' , every operator in $C \cap C'$ commutes with every operator in the symmetric difference $C \nabla C'$, so theorem 9.1 applies, and we can represent these two sets of observables as acting on different factors of a tensor product. The standard version of no-signalling can now be used to show that the marginals on the first factor are independent of the choice of measurement on the second.

9.1. Greenberger, Horne and Zeilinger models

We shall briefly review how GHZ models, which were used in section 6, can be represented in quantum mechanics. For $n > 2$, we take the Hilbert space to be the tensor product of n qubit spaces. The local observables in each factor are the X and Y spin measurements, represented in the Z basis by eigenvectors for spin Right or Left along the x -axis, with basis vectors

$$\frac{|\uparrow\rangle + |\downarrow\rangle}{\sqrt{2}}, \quad \frac{|\uparrow\rangle - |\downarrow\rangle}{\sqrt{2}}$$

and similarly for spin Forward or Back along the y -axis, with basis vectors

$$\frac{|\uparrow\rangle + i|\downarrow\rangle}{\sqrt{2}}, \quad \frac{|\uparrow\rangle - i|\downarrow\rangle}{\sqrt{2}}.$$

We shall label the outcomes as 0 for spin Right for X and spin Forward for Y and 1 for spin Left and spin Back, respectively.

The model is then generated by the GHZ state [12, 49], written in the Z basis as

$$\frac{|\uparrow \cdots \uparrow\rangle + |\downarrow \cdots \downarrow\rangle}{\sqrt{2}}.$$

If we measure each particle with a choice of X or Y observable, the probability for each outcome is given by the square modulus of the inner product

$$|\langle \text{GHZ} | b_1 \cdots b_n \rangle|^2,$$

where b_i is the basis vector corresponding to the given outcome in the i th component.

This computation is controlled by the product of the $|\downarrow\rangle$ -coefficients of the basis vectors, and hence by the cyclic group of order 4 generated by i .

The following table gives the coefficients of the $|\downarrow\rangle$ components labelled by the measurement and outcome:

	0	1
X	+1	-1
Y	+i	-i

⁹ We thank one of the journal referees for this observation.

The probability table for this model can be specified as follows.

- Each row with an odd number of Y measurements has full support.
- Each row in which the number of Y measurements is a multiple of 4 has as support those entries with an even number of outcomes labeled 1.
- Each row in which the number of Y measurements is $4k + 2$ has as support those entries with an odd number of outcomes labeled 1.
- In each case, the distribution is uniform on the support.

Note that here by ‘row’ we mean a row of the probability table, i.e. the distribution on the set of sections over a given measurement context.

Thus the interesting structure of this model arises purely from the support.

9.2. Kochen–Specker representations

We shall now discuss how the abstract discussion of Kochen–Specker situations in section 7 can be represented in terms of quantum mechanics.

We shall consider a particular form of dichotomic observables. Given unit vectors $\mathbf{e}_1, \dots, \mathbf{e}_k$ representing distinct rays in a Hilbert space \mathcal{H} , we write

$$A_{\mathbf{e}_j} := j \cdot \mathbf{P}_{\mathbf{e}_j} + 0 \cdot \mathbf{P}_{\mathbf{e}_j}^\perp.$$

Then we can take $\mathcal{X} = \{A_{\mathbf{e}_1}, \dots, A_{\mathbf{e}_k}\}$ as a set of measurements. Note that $A_{\mathbf{e}_i}$ commutes with $A_{\mathbf{e}_j}$ if and only if \mathbf{e}_i is orthogonal to \mathbf{e}_j . Also, the composition of a set of commuting observables $\{A_{\mathbf{e}_i}\}_{i \in I}$ will have a spectral decomposition of the form

$$\sum_{i \in I} i \cdot \mathbf{P}_{\mathbf{e}_i} + 0 \cdot \mathbf{P}_{\{\mathbf{e}_i | i \in I\}^\perp}.$$

If we measure any state with this observable, the outcome must be that we get exactly one of the branches $\mathbf{P}_{\mathbf{e}_i}$, with eigenvalue i , or that we get ‘none of the above’, corresponding to the branch $\mathbf{P}_{\{\mathbf{e}_i | i \in I\}^\perp}$, with eigenvalue 0. Moreover, if the cardinality of I equals the dimension of the Hilbert space, then the latter case cannot apply.

If we now consider how outcomes are represented in the sheaf \mathcal{E} , we see that we indeed have an *a priori* condition on those sections s which can be in the support of a distribution coming from a quantum state, as desired. Namely, using x_i as a label for $A_{\mathbf{e}_i}$ and taking $s(x_i) = 1$ for the outcome corresponding to $\mathbf{P}_{\mathbf{e}_i}$ for this observable, we see that the only sections which are possible are those which assign 1 to at most one measurement. Moreover, for those sets of compatible observables whose cardinality equals the dimension of the space—which must necessarily be maximal, and hence will appear in the measurement cover—exactly one measurement must be assigned 1.

Thus if we take a set of unit vectors indexed by X , such that each vector is contained in at least one orthonormal basis indexed by a subset of X , the measurement cover \mathcal{M} represented by the observables A_x will have the following key property: for *any* quantum state ρ , the support of the corresponding empirical model will satisfy the formula $\text{ONE}(C)$ for each context C in \mathcal{M} . So the problem of exhibiting a state-independent form of strong contextuality has been reduced to the problem of finding a Kochen–Specker graph, as described in section 7.

9.3. Bell-type scenarios and Kochen–Specker theorems

The measurement covers arising from Bell-type scenarios are a rather special class, which can be characterized as follows.

Proposition 9.3. *A measurement cover \mathcal{M} arises from a Bell-type scenario if and only if it is the family of maximal cliques of a graph $G = (X, E_G)$ which is the complement of an equivalence relation R on X :*

$$E_G = \{\{x, y\} \mid \neg(xRy)\}.$$

Proof. Equivalence relations are in bijective correspondence to partitions $X = \bigsqcup_i X_i$. The maximal cliques of G are exactly the *transversals* of this partition, i.e. the sets $T \subseteq X$ such that T intersects with each X_i in exactly one element. \square

Note, in particular, that in Bell-type scenarios, *incompatibility is transitive*. This is by no means the general case. In terms of operators, A and B may commute, while C may fail to commute with either.

The more complex configurations typical of Kochen–Specker constructions can never arise from Bell-type scenarios.

Proposition 9.4. *Consider a measurement cover \mathcal{M} of Bell type, and any quantum representation of \mathcal{M} . For any $s \in \mathcal{E}(C)$ with $C \in \mathcal{M}$, there is a quantum state ρ such that s is in the support of ρ_C .*

Proof. Given s , we define the local state $\rho_i := |\psi_i\rangle\langle\psi_i|$ for each i , where ψ_i is the eigenvector corresponding to the outcome specified by s for the measurement at i . Then the model defined by the state $\rho := \rho_1 \otimes \cdots \otimes \rho_n$ has s in its support. \square

Hence there is no Kochen–Specker-type theorem for Bell-type scenarios. While, as we have seen, there *are* model-specific strong contextuality results, there are no *generic* results. The measurement covers arising from these scenarios are simply not rich enough in their combinatorial structure of overlapping intersections to support a result of this form.

10. Postlude

Our treatment of non-locality and contextuality makes a number of points.

- Firstly, it is carried out at a high level of generality, and without any presupposition of quantum mechanics. None of the characteristic mathematical structures of quantum mechanics, such as complex numbers, Hilbert spaces, operator algebras or projection lattices, are needed to expose the key structural issues. This characteristic is shared to some extent by other foundational approaches, such as generalized probabilistic theories [17], but these formalisms are still rather closer to that of quantum mechanics, and indeed have been directly suggested by it. Structures such as sheaves and presheaves varying over contexts can be seen as basic elements of a general ‘logic of contextuality’, and related structures have been used for a wide range of purposes, e.g. in the semantics of computation [51–53]. This opens up the possibility of making some interesting connections between the study of non-locality and contextuality in physics, and ideas arising in other fields.

- The sheaf-theoretic language, which directly captures the idea of structures varying over contexts, is a canonical setting for studying contextuality. Moreover, as we have seen, the gluing conditions and the existence of global sections capture the essential content of non-locality and contextuality in a canonical mathematical form.

This opens the door to the use of the powerful methods of sheaf theory, which plays a major role in modern mathematics, in analyzing the structure of non-locality and contextuality. These notions are still poorly understood in multipartite and higher-dimensional settings.

In [54], the first author, with Shane Mansfield and Rui Soares Barbosa, define an Abelian presheaf based on the support of an empirical model. The Čech cohomology of this presheaf with respect to the measurement cover is used to define a cohomological obstruction to locality or non-contextuality, as a certain cohomology class. It is shown for a number of salient examples, including PR boxes, GHZ states, the Peres–Mermin square and the 18-vector configuration from [22] giving a proof of the Kochen–Specker theorem in four dimensions, that this obstruction does not vanish, thus yielding cohomological witnesses for contextuality. While these results are preliminary, they suggest that the use of cohomological methods in the study of non-locality and contextuality has some promise.

- The canonical form of description of the key concepts in terms of the existence of global sections largely replaces any explicit mention of hidden variables. These appear only in section 8, in the context of a foundational result showing the equivalence of local hidden-variable realizations to the existence of global sections. On the other hand, empirical models, which can be seen as directly related to observation and experiment, play a prominent role throughout the paper.

There is also an interesting conceptual point to be made in relation to *incompatibility of measurements*. Usually, this is taken to be a postulate of quantum mechanics, and specific to the quantum-mechanical formalism of non-commuting observables. However, in the light of general results such as those obtained in this paper, in a line of work going back to that of Fine [23], a different view emerges. The incompatibility of certain measurements can be interpreted as the impossibility—in the sense of mathematically provable non-existence—of joint distributions on *all* measurements which marginalize to yield the observed empirical distributions. Thus, if we refer to the experimental scenario with which we began section 2, this shows that there cannot be, even in principle, any such scenario in which all measurements can be carried out jointly, which is consistent with the actually observed outcomes.

Thus the incompatibility of certain measurements is revealed as a theory-independent structural impossibility result for certain families of empirical distributions. These families include those predicted by quantum mechanics, and confirmed by experiment, but the result itself is completely independent of quantum mechanics. Thus in this sense, we can say that incompatibility is *derived* rather than *assumed*.

10.1. Related work

The present paper builds on our previous work, in particular [19] by the first author, and [55, 56] by the second author (with H Jerome Keisler and Nason Yanofsky, respectively). A natural direction for generalization of the results in the present paper would be from the finite setting considered here to the measure-theoretic one studied in [55]; note that the distribution functor can be defined over general measure spaces [57].

Since we use sheaf theory as our mathematical setting, there is an obvious point of comparison with the topos approach, as developed by Isham and Butterfield, Döring and Isham and Heunen *et al* [58–60].

The general point that presheaves varying over a partially ordered set of contexts provide a natural mathematical setting for studying contextuality phenomena is certainly a common feature. It should also be mentioned that presheaves have been used for similar purposes in the context of the semantics of computation, e.g. in the Reynolds–Oles functor-category semantics for programs with state [51, 52] and in the presheaf models for concurrency of Cattani and Winskel [53].

A more specific source of inspiration is the important insight given in [58], which initiated the whole topos approach, that the Kochen–Specker theorem could be reformulated very elegantly in presheaf terms as stating the non-existence of global sections of a certain presheaf.

On the other hand, there are several differences between this paper and the topos approach. For example, the topos approach focuses on a specific structure, the spectral presheaf, based on an operator algebra. In this sense, it uses concepts specific to quantum mechanics from the very start. Moreover, many of the key structures introduced in our paper, such as the distribution functor and measurement covers, do not appear in the topos approach. One of our central objectives is to give a unified account of contextuality and non-locality, but locality issues have not been considered in the topos approach; nor has extendability, another key topic for us. It will, of course, be interesting to see if additional commonalities develop in future work.

The appearance of negative probabilities in quantum mechanics has a long history. The Wigner quasi-probability distribution [14], further developed by Moyal [16], is a phase-space representation of quantum mechanics using negative probabilities. Feynman views such negative probabilities as a calculational convenience [17]. He explains that the appearance of a negative probability for a certain outcome does not invalidate the theory being used. Rather, this tells us that the relevant conditions cannot be realized or that the outcome cannot be verified or both. More specifically related to what we do, Sudarshan and Rothman [61] show that a local hidden-variable analysis of the Bell model is possible if certain values of the hidden variable arise with negative probability. Finally, in Dirac [15], negative probabilities enter in the relativistic extension of quantum mechanics.

Acknowledgments

We thank Noson Yanofsky for many important discussions and for encouragement during the development of the ideas in this paper. Harvey Brown, Andreas Döring, Peter Johnstone, Aleks Kissinger, Ray Lal, Pierfrancesco La Mura, Elliot Lipnowski, Shane Mansfield, John Myers, Andrei Savochkin, Natalya Vinokurova and audiences at several meetings, including the 2010–2011 Clifford Lectures, Tulane University, March 2011, the LINT Workshop in Oxford, April 2011, the McGill University Bellairs Workshop, April 2011, The Knots in Washington Workshop, George Washington University, May 2011, the Workshop on Topological Quantum Information, Pisa, May 2011, and the Conference on Unconventional Computation, Turku, June 2011, provided valuable input. Financial support from the Stern School of Business, the CUNY Graduate Center, EPSRC Senior Research Fellowship EP/E052819/1 and the US Office of Naval Research grant number N000141010357 is gratefully acknowledged. We thank the two anonymous journal referees for their comments, which led to a number of clarifications and to improved references to the literature. One of the referees, in particular, made a number of

detailed and perceptive comments, which led to several significant improvements to the revised version of the paper.

Appendix

Firstly, we recall some set-theoretic notations.

We write $|S|$ for the cardinality of a set S . If $f : X \rightarrow Y$ is a function and $X' \subseteq X$, we write $f|_{X'} : X' \rightarrow Y$ for the restriction of f to X' . We write Y^X for the set of functions from X to Y , and $\mathcal{P}(X)$ for the powerset of X .

A family of sets $\{X_i\}_{i \in I}$ is *disjoint* if $X_i \cap X_j = \emptyset$ whenever $i \neq j$. We write $\coprod_{i \in I} X_i$ for the union of a disjoint family. Given a disjoint family $\{X_i\}_{i \in I}$, there is an isomorphism

$$\mathcal{P}\left(\coprod_{i \in I} X_i\right) \xrightarrow{\cong} \prod_{i \in I} \mathcal{P}(X_i) :: S \mapsto (S \cap X_i)_{i \in I}.$$

A *category* has a collection of objects A, B, C, \dots and arrows f, g, h, \dots . Each arrow has specified *domain* and *codomain* objects: the notation is $f : A \rightarrow B$ for an arrow f with domain A and codomain B . Given arrows $f : A \rightarrow B$ and $g : B \rightarrow C$, we can form the *composition* $g \circ f : A \rightarrow C$. Composition is associative, and there are identity arrows $\text{id}_A : A \rightarrow A$ for each object A , with $f \circ \text{id}_A = f$, $\text{id}_B \circ g = g$, for every $f : A \rightarrow B$ and $g : C \rightarrow A$.

Our main examples of categories will be **Set**, with sets as objects and with functions as arrows; and partially ordered sets (P, \leq) , where there is a single arrow from p to q if $p \leq q$ and none otherwise. The *opposite category* P^{op} is the category formed from the opposite poset (P, \geq) .

If \mathcal{C} and \mathcal{D} are categories, a *functor* $F : \mathcal{C} \rightarrow \mathcal{D}$ assigns an object FA of \mathcal{D} to each object A of \mathcal{C} ; and an arrow $Ff : FA \rightarrow FB$ of \mathcal{D} to every arrow $f : A \rightarrow B$ of \mathcal{C} . These assignments must preserve composition and identities: $F(g \circ f) = F(g) \circ F(f)$, and $F(\text{id}_A) = \text{id}_{FA}$.

A *presheaf* on a poset P is a functor $P^{\text{op}} \rightarrow \mathbf{Set}$.

References

- [1] Bell J S 1964 On the Einstein–Podolsky–Rosen paradox *Physics* **1** 195–200
- [2] Kochen S and Specker E P 1967 The problem of hidden variables in quantum mechanics *J. Math. Mech.* **17** 59–87
- [3] Bartosik H, Klepp J, Schmitzer C, Sponar S, Cabello A, Rauch H and Hasegawa Y 2009 Experimental test of quantum contextuality in neutron interferometry *Phys. Rev. Lett.* **103** 40403
- [4] Kirchmair G, Zähringer F, Gerritsma R, Kleinmann M, Gühne O, Cabello A, Blatt R and Roos C F 2009 State-independent experimental test of quantum contextuality *Nature* **460** 494–7
- [5] Heywood P and Redhead M L G 1983 Nonlocality and the Kochen–Specker paradox *Found. Phys.* **13** 481–99
- [6] Stairs A 1983 Quantum logic, realism and value definiteness *Phil. Sci.* **50** 578–602
- [7] Mermin N D 1990 Simple unified form for the major no-hidden-variables theorems *Phys. Rev. Lett.* **65** 3373–6
- [8] Cabello A, Severini S and Winter A 2010 Non-contextuality of physical theories as an axiom, arXiv:1010.2163
- [9] Mac Lane S and Moerdijk I 1992 *Sheaves in Geometry and Logic: a First Introduction to Topos Theory* (Berlin: Springer)
- [10] Popescu S and Rohrlich D 1994 Quantum nonlocality as an axiom *Found. Phys.* **24** 379–85
- [11] Hardy L 1993 Nonlocality for two particles without inequalities for almost all entangled states *Phys. Rev. Lett.* **71** 1665–8

- [12] Greenberger D M, Horne M A and Zeilinger A 1989 Going beyond Bell's theorem *Bell's Theorem, Quantum Theory and Conceptions of the Universe* ed M Kafatos (Dordrecht: Kluwer) pp 69–72
- [13] Lal R 2011 private communication
- [14] Wigner E 1932 On the quantum correction for thermodynamic equilibrium *Phys. Rev.* **40** 749
- [15] Dirac P A M 1942 The physical interpretation of quantum mechanics *Proc. R. Soc. A* **180** 1–40
- [16] Moyal J E 1949 Quantum mechanics as a statistical theory *Math. Proc. Camb. Phil. Soc.* **45** 99–124
- [17] Feynman R P 1987 Negative probability *Quantum Implications: Essays in Honour of David Bohm* ed B J Hiley and F D Peat (London: Routledge) pp 235–48
- [18] Ghirardi G C, Rimini A and Weber T 1980 A general argument against superluminal transmission through the quantum mechanical measurement process *Lett. Nuovo Cimento (1971–85)* **27** 293–8
- [19] Abramsky S 2010 Relational Hidden Variables and non-locality *Studia Logica* at press (arXiv:1007.2754)
- [20] Fritz T 2009 Possibilistic Physics *FQXi Essay* <http://www.fqxi.org/community/forum/topic/569>
- [21] Jacobs B 2010 Convexity, duality and effects *Theoretical Computer Science—6th IFIP TC 1/WG 2.2 International Conference, TCS 2010* **323** 1–19
- [22] Cabello A, Estebaranz J M and García-Alcaine G 1996 Bell–Kochen–Specker theorem: a proof with 18 vectors *Phys. Lett. A* **212** 183–7
- [23] Fine A 1982 Hidden variables, joint probability and the Bell inequalities *Phys. Rev. Lett.* **48** 291–5
- [24] Basoalto R M and Percival I C 2003 Bell Test and CHSH experiments with more than two settings *J. Phys. A: Math. Gen.* **36** 7411
- [25] Pironio S 2005 Lifting Bell inequalities *J. Math. Phys.* **46** 062112
- [26] Mansfield S and Fritz T 2011 Hardy's Non-locality Paradox and possibilistic conditions for non-locality arXiv:1105.1819v1
- [27] Feynman R P 1982 Simulating physics with computers *Int. J. Theor. Phys.* **21** 467–88
- [28] Peres A 1990 Incompatible results of quantum measurements *Phys. Lett. A* **151** 107–8
- [29] Mermin N D 1990 Quantum mysteries revisited *Am. J. Phys.* **58** 731–4
- [30] Elitzur A C, Popescu S and Rohrlich D 1992 Quantum nonlocality for each pair in an ensemble *Phys. Lett. A* **162** 25–8
- [31] Barrett J, Kent A and Pironio S 2006 Maximally nonlocal and monogamous quantum correlations *Phys. Rev. Lett.* **97** 170409
- [32] Aolita L, Gallego R, Acín A, Chiuri A, Vallone G, Mataloni P and Cabello A 2011 Fully nonlocal quantum correlations arXiv:1105.3598
- [33] Montanari U 1974 Networks of constraints: fundamental properties and applications to picture processing *Inf. Sci.* **7** 95–132
- [34] Kumar V 1992 Algorithms for constraint-satisfaction problems: a survey *AI Mag.* **13** 32
- [35] Waegell M and Aravind P K 2011 Parity proofs of the Kochen–Specker theorem based on the 24 rays of Peres arXiv:1103.6058v1
- [36] Liang Y C, Spekkens R W and Wiseman H M 2010 Specker's parable of the over-protective seer: implications for contextuality, nonlocality and complementarity arXiv:1010.1273
- [37] Berge C and Duchet P 1984 Strongly perfect graphs *Top. Perfect Graphs* **21** 57–61
- [38] Lovász L, Saks M and Schrijver A 1989 Orthogonal representations and connectivity of graphs *Linear Algebra Appl.* **114** 439–54
- [39] Peres A 1991 Two simple proofs of the Kochen–Specker theorem *J. Phys. A: Math. Gen.* **24** L175
- [40] Bub J 1996 Schütte's tautology and the Kochen–Specker theorem *Found. Phys.* **26** 787–06
- [41] Pavičić M, Merlet J P, McKay B and Megill N D 2005 Kochen–Specker vectors *J. Phys. A: Math. Gen.* **38** 1577
- [42] Lovász L, Saks M and Schrijver A 2000 A correction: orthogonal representations and connectivity of graphs *Linear Algebra Appl.* **313** 101–6
- [43] Lovász L 1979 On the Shannon capacity of a graph *Inf. Theory IEEE Trans.* **25** 1–7
- [44] Dickson W M 1999 *Quantum Chance and Non-Locality* (Cambridge: Cambridge University Press)

- [45] Jarrett J P 1984 On the physical significance of the locality conditions in the Bell arguments *Noûs* **18** 569–89
- [46] Shimony A 1986 Events and processes in the quantum world *Quantum Concepts in Space and Time* (Oxford: Clarendon) pp 182–203
- [47] Tsirelson B S 2006 Bell inequalities and operator algebras [private communication](#)
- [48] Scholz V B and Werner R F 2008 Tsirelson’s problem arXiv:0812.4305
- [49] Greenberger D M, Horne M A, Shimony A and Zeilinger A 1990 Bell’s theorem without inequalities *Am. J. Phys.* **58** 1131
- [50] Barrett J 2007 Information processing in generalized probabilistic theories *Phys. Rev. A* **75** 032304
- [51] Oles F J 1985 Type categories, functor categories and block structure *Algebraic Semantics* ed M Nivat and J C Reynolds (Cambridge: Cambridge University Press) pp 543–74
- [52] Tennent R 1986 Functor-category semantics of programming languages and logics *Category Theory and Computer Programming (Lecture Notes in Computer Science vol 240)* ed D Pitt, S Abramsky, A Poigné and D Rydeheard (Berlin: Springer) pp 206–24
- [53] Cattani G and Winskel G 1997 Presheaf models for concurrency *Computer Science Logic* (Berlin: Springer) pp 58–75
- [54] Abramsky S, Barbosa R S and Mansfield S 2011 The cohomology of nonlocality and contextuality *Electronic Proceedings in Theoretical Computer Science* at press
- [55] Brandenburger A and Keisler H J 2011 *What does a hidden variable look like?* <http://pages.stern.nyu.edu/~abranden/wdhv-03-03-11.pdf>
- [56] Brandenburger A and Yanofsky N 2008 A classification of hidden-variable properties *J. Phys. A: Math. Theor.* **41** 425302
- [57] Giry M 1982 A categorical approach to probability theory *Categorical Aspects of Topology and Analysis (Lecture Notes in Mathematics vol 915)* (Berlin: Springer) pp 68–85
- [58] Isham C J and Butterfield J 1998 Topos perspective on the Kochen–Specker theorem: I. Quantum states as generalized valuations *Int. J. Theor. Phys.* **37** 2669–733
- [59] Döring A and Isham C J 2008 A topos foundation for theories of physics: II. Daseinisation and the liberation of quantum theory *J. Math. Phys.* **49** 053516
- [60] Heunen C, Landsman N P and Spitters B 2009 A topos for algebraic quantum theory *Commun. Math. Phys.* **291** 63–110
- [61] Sudarshan E C G and Rothman T 1993 A new interpretation of Bell’s inequalities *Int. J. Theor. Phys.* **32** 1077–86