A Unified Sheaf-Theoretic Account Of Non-Locality and Contextuality*

Samson Abramsky† and Adam Brandenburger‡

February 1, 2011

Abstract

A number of landmark results in the foundations of quantum mechanics show that quantum systems exhibit behaviour that defies explanation in classical terms, and that cannot be accounted for in such terms even by postulating "hidden variables" as additional unobserved factors. Much has been written on these matters, but there is surprisingly little unanimity even on basic definitions or the inter-relationships among the various concepts and results. We use the mathematical language of sheaves and monads to give a very general and mathematically robust description of the behaviour of systems in which one or more measurements can be selected, and one or more outcomes observed. We say that an empirical model is extendable if it can be extended consistently to all sets of measurements, regardless of compatibility. A hidden-variable model is factorizable if, for each value of the hidden variable, it factors as a product of distributions on the basic measurements. We prove that an empirical model is extendable if and only if there is a factorizable hidden-variable model which realizes it. From this we are able to prove generalized versions of well-known No-Go theorems. At the conceptual level, our equivalence result says that the existence of incompatible measurements is the essential ingredient in non-local and contextual behavior in quantum mechanics.

1 Introduction

An experiment consists of making one or more measurements on a system of interest, and recording the outcomes of the measurements. For each set of measurements that can be chosen, there will be a frequency distribution of the various joint outcomes obtained under that set of measurements. A system might be made up of different parts, where Alice performs measurements on one part, Bob on another part, and so on. An important case is when the different parts of the system are spatially separated.

Sometimes, there are constraints on what sets of measurements can be simultaneously undertaken on a system, because making one measurement 'interferes' with making a second. Such situations arise most famously in the quantum world, where various measurements (e.g., position and momentum) are incompatible. Quantum mechanics is the primary motivation for this paper, so our framework should allow for incompatibilities among measurements. One can also imagine such situations arising in the classical world.

*We thank Noson Yanofsky for many important discussions and for encouragement throughout the development of the ideas in this paper. Harvey Brown, Andreas Döring, Pierfrancesco La Mura, and Elliot Lipnowski provided valuable input. Financial support from the Stern School of Business, the CUNY Graduate Center, EPSRC Senior Research Fellowship EP/E052819/1, and the U.S. Office of Naval Research Grant Number N000141010357 is gratefully acknowledged.

†Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, U.K., samson@comlab.ox.ac.uk, http://web.comlab.ox.ac.uk/people/samson.abramsky.

‡Stern School of Business, New York University, 44 West Fourth Street, New York, NY 10012, adam.brandenburger@stern.nyu.edu, www.stern.nyu.edu/∼abranden.

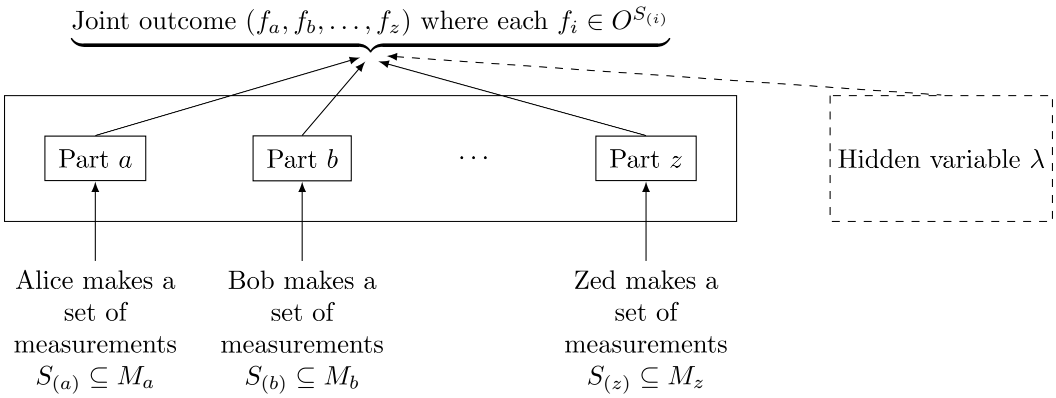
The mathematical language we will use to describe experiments is sheaf theory. This allows us to give a very general and flexible treatment. We begin this approach in Section 2. Firstly, we will give a heuristic account of some of the main ideas of the paper. In this heuristic discussion, we will speak of probability distributions specifically. However, the actual mathematical scope of the paper will be much broader, and probability distributions will appear as a very special case of the general theory.

1.1 Empirical and Hidden-Variable Models

A **system** will be made up of a number of spatially separated **parts**. Associated with each part i will be a set M_i of basic **measurements** that can be performed on that part. Associated with each measurement is a set O of possible **outcomes**. (By suitable expansion, we can take the outcome set to be common across all measurements.) Next is the idea of a **compatibility structure**. At each part there will be a family $\mathcal{C}_i$ of subsets of M_i specifying which measurements are compatible, i.e., can be performed together as a single joint measurement. Take the M_i to be disjoint, and form the disjoint union $M = \coprod_{i \in I} M_i$. We can then form a compatibility structure $\mathcal{C}$ on M, given by:

$$\mathcal{C} = \{S \subseteq M \mid \forall i \in I.\, S_{(i)} \in \mathcal{C}_i\},$$

where $S_{(i)} = S \cap M_i$. Thus, we take measurements performed in different parts of the system to be automatically compatible, but require measurements within a part to respect that part's compatibility conditions. The following picture shows our formalization of an experiment — or, what we shall from now on call an **empirical model**. (For now, ignore the box labelled "Hidden variable λ.")



Two special cases are immediate. In the first case, each $\mathcal{C}_i$ consists of the singletons $\{m_i\}$ alone. Thus, a joint measurement consists of at most one measurement on each part. This is the setting for studying non-locality in the sense of Bell [4], where the question is how to explain correlations that arise across outcomes at different parts. A second special case is when there is just one part, and the compatibility structure $\mathcal{C}$ consists of more than the singletons but less than the power set. This is a starting point for studying contextuality in the sense of Kochen-Specker [26], where the question is whether incompatible sets of measurements can be put together in a certain sense.

Empirical models can be embellished by adding to them a set, customarily denoted Λ, of hidden (unobserved) variables. Taking into account the box labelled "Hidden variable λ" in Figure 1, we obtain the picture of a **hidden-variable model**. We see that the hidden variable λ interacts with the system to determine (probabilistically) the outcomes of measurements.

A hidden-variable model **realizes** an empirical model if, for each set of compatible measurements, the probability distribution on outcomes that is induced by the hidden-variable model — when we average over the possible values of λ — agrees with the probability distribution specified by the empirical model. That is, as far as observable variables are concerned, the hidden-variable model behaves the same way as the empirical model. There are various properties one can ask of a hidden-variable model, such as determinism and factorizability. (We describe these properties below.) If an empirical model can be realized by a hidden-variable model satisfying properties of this kind, then this provides a particular kind of explanation of the actual data.

1.2 Overview of Results

We consider several properties of empirical models:

- **No Signalling** A generalized version of No-Signalling emerges as a naturality property in our framework. This generalized form applies to any set of compatible measurements, whether or not they are regarded as spatially distributed. We have not found an explicit discussion of this point in the literature, although this generalized version is easily seen to be valid in quantum mechanics.
- **Non-Contextuality** This property comes in two forms. We say an empirical model is contextual if it cannot be realized by a factorizable (see below) hidden-variable model. It is strongly contextual if there is no assignment of outcomes to measurements — no function from M to O — whose restriction to each set of compatible measurements lies in the support of the associated probability measure on outcomes. Strong contextuality is strictly stronger than contextuality.
- **Extendability** This asks for the existence of an extension of the empirical model to one defined on all subsets of measurements, regardless of compatibility. This is a non-trivial property because of the naturality requirement on models, which implies in particular that the marginals of the extended model must agree with the distributions assigned by the original model.

We also consider several properties of hidden-variable models:

- **Parameter Independence** This emerges as a naturality property in our framework.
- **λ-Independence** This corresponds to the assumption that the space of hidden variables is a **constant presheaf**. Without this assumption, it is trivially possible to introduce deterministic hidden variables.
- **Factorizability** This says that, given the value of the hidden variable λ, the probability of a joint outcome for a set of measurements can be factored into a product of probabilities. When each part's compatibility structure $\mathcal{C}_i$ consists of the singletons $\{m_i\}$, this specializes to the usual Locality property.
- **Determinism** This says that, given a set of compatible measurements and the value of λ, there is a unique joint outcome.

An easy implication is: Determinism $\Rightarrow$ Factorizability.

Our key result relating empirical and hidden-variable models can be stated as follows:

> **An empirical model is extendable if and only if it can be realized by a factorizable hidden-variable model.**

Equivalently:

> **An empirical model is extendable if and only if it is non-contextual.**

This result has many consequences, and has the overall effect of streamlining the theory considerably. To a large extent, it makes it possible to work directly with empirical models, without needing to consider hidden variables explicitly.

We use our approach to derive general forms of two types of No-Go theorems:

- Firstly, we can prove **model specific** results, essentially versions of Bell's theorem. We give a very simple and direct proof of a Hardy-style impossibility result (Hardy [19]) by showing that the empirical model in question is not extendable. This is a 'hidden-variable-free' proof, in contrast to the standard arguments.

  We show that a standard quantum system giving rise to a Hardy model, while contextual, is not strongly contextual; the support does have a global section. By contrast, GHZ models [18, 17] are strongly contextual.

- Secondly, we consider **model-independent** results. We formulate a generalized version of the Kochen-Specker theorem [26], which establishes the strong contextuality of a whole class of empirical models, by showing that a simply defined presheaf which contains the support of these models has no global sections. In particular, in the quantum case this shows that there are finite sets of measurements under which **all** quantum states, whether entangled or not, exhibit strongly contextual behaviour. We also point out some interesting connections to combinatorics and computational complexity, including a purely graph-theoretic formulation of "Kochen-Specker graphs".

We end this preview by emphasizing the conceptual significance of our result relating empirical and hidden-variable models. This tells us that the existence of incompatible measurements is an essential ingredient — arguably, **the** essential ingredient — in non-local and contextual behavior in quantum mechanics. To the best of our knowledge, the first formal result of this kind appeared in [15]. The generality of our framework allows us to produce some particularly strong and compelling evidence on this point.

1.3 Related Work

The present paper builds on our previous work, in particular [1] by the first author, and [7, 8] by the second author (with H. Jerome Keisler and Noson Yanofsky, respectively).

Since we use sheaf theory as our mathematical setting, there is an obvious point of comparison with the topos approach, as developed by Isham, Butterfield, Döring, Heunen, Landsman, Spitters et al. [21, 13, 20].

There are two aspects where we see a clear connection:

- The general point that presheaves varying over a poset of contexts provides 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 [36, 41], and in the presheaf models for concurrency of Cattani and Winskel [11].

- More specifically, our development of a generalized Kochen-Specker theorem in Section 12 draws on the important insight in [21], 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. We relate our approach to the formulation in [21] in Section 12.4.

On the other hand, there are many differences in emphasis, objectives, and technical development between the present work and the topos approach. For example, sheaves and monads feature prominently in the present paper, but have not appeared in the topos approach to date, as far as we are aware. 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. In the other direction, the ambient presheaf topos and its internal logic, which is central to the topos approach, does not play a rôle in our work.

2 Mathematical Background

The mathematical language we shall use is **sheaf theory**. There are three parameters in our approach, which allow for a considerable degree of generality and flexibility:

- The notion of **context** is parameterized by the **poset** over which the sheaves and presheaves are defined. This allows locality, contextuality, and combinations of the two to be covered in a uniform fashion.
- The notion of **behaviour** or **event** is parameterized as a **sheaf** over the poset of contexts.
- The notion of **weight** assigned by models to events is parameterized by a **commutative monad** [27], which can be regarded as expressing a general notion of **effect**, as in [34]. In particular, any commutative semiring gives rise to a **distribution monad**. Thus both probabilistic and relational models are covered, as very special cases of the general theory.

In this Section, we shall review the mathematical background assumed in this paper, and establish the notational conventions which we will use.

2.1 Notational Preliminaries

We write $|S|$ for the cardinality of a set S. If $f : X \to Y$ is a function and $X' \subseteq X$, we write $f|X' : X' \to Y$ for the restriction of f to X'. We write Y^X for the set of functions from X to Y.

A family of sets $\{X_i\}_{i\in I}$ is **disjoint** if $X_i \cap X_j = \varnothing$ whenever $i \neq j$. We write $\coprod_{i\in I} X_i$ for the union of a disjoint family. Given a family of functions $\{f_i : X_i \to Y\}_{i\in I}$, there is a unique function $f : \coprod_{i\in I} X_i \to Y$ with $f|X_i = f_i$ for all $i \in I$.

Given $S \subseteq \coprod_{i\in I} X_i$, we write $S_{(i)} = S \cap X_i$. Note that the map $S \mapsto (S_{(i)})_{i\in I}$ is an isomorphism

$$\mathcal{P}(\coprod_{i\in I} X_i) \xrightarrow{\cong} \prod_{i\in I} \mathcal{P}(X_i). \tag{1}$$

2.2 Categorical Background

We will assume some familiarity with the very basic elements of the language of category theory: categories, functors, natural transformations, products, isomorphisms, cartesian closed categories. A gentle and succinct introduction such as [39] or [2] covers all that we will need, while [3] provides a more substantial reference.

Our discussion of sheaves and presheaves will be essentially self-contained. A standard reference is [31].

We shall make some use of **commutative monads** [27], which we briefly review.

A **monad** (T, η, μ) on a category $\mathcal{C}$ is given by an endofunctor $T : \mathcal{C} \to \mathcal{C}$, and natural transformations $\eta : \mathsf{Id} \stackrel{.}{\longrightarrow} T$, $\mu : T^2 \stackrel{.}{\longrightarrow} T$, satisfying the following commutative diagrams:

$$\begin{tikzcd}
TTTX \arrow[r, "T\mu_X"] \arrow[d, "\mu_{TX}"'] & TTX \arrow[d, "\mu_X"] \\
TTX \arrow[r, "\mu_X"'] & TX
\end{tikzcd}
\qquad
\begin{tikzcd}
TX \arrow[r, "\eta_{TX}"] \arrow[dr, "1_{TX}"'] & TTX \arrow[d, "\mu_X"] & TX \arrow[l, "T\eta_X"'] \arrow[dl, "1_{TX}"] \\
& TX &
\end{tikzcd}$$

Example The powerset operation defines a covariant functor on **Set** by direct image:

$$f : X \to Y \quad \mapsto \quad \mathcal{P}f : \mathcal{P}X \to \mathcal{P}Y :: S \mapsto \{f(x) \mid x \in S\}.$$

We can define natural transformations

$$\eta_X : X \to \mathcal{P}X :: x \mapsto \{x\}, \qquad \mu_X : \mathcal{P}\mathcal{P}X \to \mathcal{P}X :: \Theta \mapsto \bigcup \Theta,$$

and $(\mathcal{P}, \eta, \mu)$ forms a monad.

Now assume that $\mathcal{C}$ has finite products. A **commutative monad** (T, η, μ, κ) additionally has a natural transformation

$$\kappa_{X,Y} : TX \times TY \to T(X \times Y)$$

which satisfies some coherence conditions:

$$\begin{tikzcd}
X \times Y \arrow[r, "\eta_X \times \eta_Y"] \arrow[dr, "\eta_{X \times Y}"'] & TX \times TY \arrow[d, "\kappa_{X,Y}"] \\
& T(X \times Y)
\end{tikzcd}
\qquad
\begin{tikzcd}
T^2X \times T^2Y \arrow[r, "\mu_X \times \mu_Y"] \arrow[d, "T\kappa_{X,Y} \circ \kappa_{TX,TY}"'] & TX \times TY \arrow[d, "\kappa_{X,Y}"] \\
T^2(X \times Y) \arrow[r, "\mu_{X \times Y}"'] & T(X \times Y)
\end{tikzcd}$$

$$\begin{tikzcd}
TX \times (TY \times TZ) \arrow[r, "\alpha_{TX,TY,TZ}"] \arrow[d, "1 \times \kappa_{TY,TZ}"'] & (TX \times TY) \times TZ \arrow[d, "\kappa_{TX,TY} \times 1"] \\
TX \times T(Y \times Z) \arrow[d, "\kappa_{X,Y \times Z}"'] & T(X \times Y) \times TZ \arrow[d, "\kappa_{X \times Y,Z}"] \\
T(X \times (Y \times Z)) \arrow[r, "T\alpha_{X,Y,Z}"'] & T((X \times Y) \times Z)
\end{tikzcd}
\qquad
\begin{tikzcd}
TX & T(X \times \mathbf{1}) \arrow[l, "T\rho_X"'] \\
TX \times \mathbf{1} \arrow[u, "\rho_{TX}"] \arrow[r, "1 \times \eta_{\mathbf{1}}"'] & TX \times T\mathbf{1} \arrow[u, "\kappa_{X,\mathbf{1}}"']
\end{tikzcd}$$

$$\begin{tikzcd}
TX \times TY \arrow[r, "\kappa_{X,Y}"] \arrow[d, "\sigma_{TX,TY}"'] & T(X \times Y) \arrow[d, "T\sigma_{X,Y}"] \\
TY \times TX \arrow[r, "\kappa_{Y,X}"'] & T(Y \times X)
\end{tikzcd}$$

where α, ρ, σ are the natural isomorphisms witnessing the associativity, unit, and symmetry properties of the product.

In our example, the powerset is a commutative monad, with the natural transformation

$$\kappa_{X,Y} : \mathcal{P}X \times \mathcal{P}Y \to \mathcal{P}(X \times Y) :: (S, T) \mapsto S \times T.$$

A commutative monad is **affine** [28] if the following diagram always commutes:

$$\begin{tikzcd}
TX \times TY \arrow[r, "\kappa_{X,Y}"] \arrow[dr, "1"'] & T(X \times Y) \arrow[d, "T\pi_1 \times T\pi_2"] \\
& TX \times TY
\end{tikzcd}$$

Proposition 2.1 ([28, 22]) *A commutative monad T is affine if and only if $\eta_{\mathbf{1}} : \mathbf{1} \to T\mathbf{1}$ is an isomorphism.*

In our example, the powerset monad is **not** affine; however, the non-empty powerset monad is.

2.3 The Distribution Monad

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:

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

Our two main examples of semirings will be the non-negative reals:

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

and the booleans:

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

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

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

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 R, this is the set of probability distributions of finite support on X; in the case of B, it is the set of non-empty finite subsets of X.

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

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

This is easily seen to be functorial:

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

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

We can also define

$$\eta_X : X \to \mathcal{D}_R(X) :: x \mapsto \delta_x$$

where $\delta_x(x) = 1$, and $\delta_x(y) = 0$, $x \neq y$. Similarly, we can define

$$\mu_X : \mathcal{D}_R(\mathcal{D}_R(X)) \to \mathcal{D}_R(X) :: \Theta \mapsto [x \mapsto \sum_{d \in \mathcal{D}_R(X)} \Theta(d) \cdot d(x)],$$

and

$$\kappa_{X,Y} : \mathcal{D}_R(X) \times \mathcal{D}_R(Y) \to \mathcal{D}_R(X \times Y) :: (d, d') \mapsto [(x, y) \mapsto d(x) \cdot d'(y)].$$

It can be verified that these are natural transformations, and that $(\mathcal{D}_R, \eta, \mu, \kappa)$ forms a commutative monad on $\mathbf{Set}$. Moreover, this monad is always affine. We verify that $\mathcal{D}_R(\pi_1) \circ \kappa = \pi_1$:

$$\mathcal{D}_R(\pi) \circ \kappa(d_1, d_2)(x) = \sum_y d_1(x) \cdot d_2(y) = d_1(x) \cdot \sum_y d_2(y) = d_1(x) \cdot 1 = d_1(x).$$

2.4 Presheaves on a poset

Let $\mathbf{P}$ be a poset. A presheaf on $\mathbf{P}$ is a functor $F : \mathbf{P}^{\mathsf{op}} \to \mathbf{Set}$. That is, we regard $\mathbf{P}$ as a category, with an arrow $\iota_{p,p'} : p \to p'$ whenever $p \leq p'$. Now F assigns a set $F(p)$ to each $p \in \mathbf{P}$, and a function $F(\iota_{p,p'}) : F(p') \to F(p)$ when $p \leq p'$. These assignments satisfy:

$$F(\iota_{p,p}) = \mathsf{id}_{F(p)}, \qquad F(\iota_{p,p''}) = F(\iota_{p',p''}) \circ F(\iota_{p.p'}), \quad p \leq p' \leq p''.$$

We refer to the elements of $F(p)$ as **sections**, and write $s{\restriction}p := F(\iota_{p,p'})(s)$ when $s \in F(p')$ and $p \leq p'$.

A morphism of presheaves on $\mathbf{P}$ is a natural transformation $t : F \stackrel{\cdot}{\longrightarrow} G$. Thus for each $p \in \mathbf{P}$, there is a function $t_p : F(p) \to G(p)$, such that, whenever $p \leq p'$:

$$t_p(s{\restriction}p) = t_{p'}(s){\restriction}p, \qquad s \in F(p').$$

Presheaves on $\mathbf{P}$ and their morphisms form a category, which we denote by $\mathbf{Set}^{\mathbf{P}^{\mathrm{op}}}$ or $\hat{\mathbf{P}}$.

A commutative monad (T, η, μ, κ) on $\mathbf{Set}$ lifts pointwise to a commutative monad on $\mathbf{Set}^{\mathbf{P}^{\mathrm{op}}}$. Given a presheaf F, $T(F)$ is defined by composition:

$$T(F)(X) := T(F(X)).$$

Similarly for the natural transformations η, μ, κ.

2.5 Sheaves

We now suppose that our poset $\mathbf{P}$ is **bounded complete**, meaning that every subset with an upper bound has a least upper bound, also referred to as a join. This implies that it also has all non-empty meets. We shall write the least upper bound or join of a bounded set $\{p_j\}_{j \in J}$ as $\bigvee_{j \in J} p_j$. The least upper bound of the empty set is the least element of the poset, which we write as 0.

We say that a presheaf F on $\mathbf{P}$ is a **sheaf** if it satisfies the following condition:

Whenever $p = \bigvee_{j \in J} p_j$, and there are $s_j \in F(p_j)$ for each $j \in J$, such that the **compatibility** condition

$$s_j{\restriction}(p_j \wedge p_k) = s_k{\restriction}(p_j \wedge p_k)$$

holds for all $j, k \in J$, then there exists a unique $s \in F(p)$ with $s{\restriction}p_j = s_j$ for all $j \in J$.

The sheaf condition has a standard alternative formulation as requiring the exactness of a certain diagram. We note a particular case of this. We say that a family $\{p_j\}_{j \in J}$ is a **disjoint cover** of p if $p = \bigvee_{j \in J} p_j$, and $p_j \wedge p_k = 0$ for all $j \neq k$.

Proposition 2.2 *Suppose that $\{p_j\}_{j \in J}$ is a disjoint cover of p, and F is a sheaf. Then $F(p)$ is the product in* **Set** *of $\{F(p_j)\}_{j \in J}$.*

Proof Firstly, the restriction maps provide projections:

$$\pi_j : F(p) \to F(p_j) :: s \mapsto s{\restriction}p_j.$$

Now suppose we are given a disjoint cover $\{p_j\}_{j \in J}$ of p, a set X, and functions $f_j : X \to F(p_j)$, $j \in J$. For each $x \in X$, we have $s_j := f_j(x) \in F(p_j)$. Since the empty family covers 0, the sheaf condition implies that $F(0) = \mathbf{1}$, and thus the family $\{s_j\}$ satisfies the compatibility conditions. Hence the sheaf condition for F implies that there is a unique $s \in F(p)$ such that $\pi_j(s) = s{\restriction}p_j = s_j$. Applying this to each element of X, we obtain a unique map $f : X \to F(p)$ fulfilling the universal property of the product. $\square$

Remark A bounded complete poset fails to be a complete lattice only in that it lacks a top element. This will turn out to be a critical point in our discussion of contextuality, locality and No-Go theorems. The existence of a top element, corresponding to the idea that all measurements are compatible, will entail extendability, and hence factorizability, with locality as a special case. Moreover, a converse will also hold: see the E = F Theorem 8.4. Thus, as discussed in Section 1.2, incompatibility emerges as the key to non-local and contextual behavior in quantum mechanics.

3 Contexts and Events

3.1 Contexts

We shall view the poset $\mathbf{P}$ as articulating the structure of a set of **contexts**. In particular, we shall be interested in posets of the following kind. Fix a set M of basic measurements. A **compatibility structure** $\mathcal{C}$ on M is a family of subsets of M such that:

- $\bigcup \mathcal{C} = M$.
- $\mathcal{C}$ is **downwards closed**, *i.e.* if $S \subseteq S'$ and $S' \in \mathcal{C}$, then $S \in \mathcal{C}$.

The reading of $\mathcal{C}$ is that it expresses which measurements are **compatible**, *i.e.* can be performed together as a single joint measurement. The downwards closure condition expresses the idea that a subset of a compatible family is compatible. The natural order on this structure is subset inclusion. We refer to this poset as $\mathbf{P}_{\mathcal{C}}$. It provides a setting for studying **contextuality**.

An important point in our approach is that this study of compatibility structures subsumes the study of spatially distributed, multipartite systems. Such systems give rise to compatibility structures in a natural fashion:

- Consider a disjoint family $\{M_i\}_{i \in I}$. We think of I as labelling the parts of a system, which may be space-like separated; M_i is the set of basic measurements which can be performed at part i. We form the disjoint union M of this family. A **global measurement** is a subset of M containing at most one element of M_i for each $i \in I$. Thus we regard measurements performed in different parts of the system as compatible, but do not allow for compatible measurements in the same part. The set of subsets of this form, ordered by inclusion, forms a compatibility structure $\mathcal{C}_{\mathsf{L}}$. It provides a setting for studying **locality**.

- We can also consider combinations of local compatibility and distributed structure. A **distributed compatibility structure** is a family $\{(M_i, \mathcal{C}_i)\}_{i \in I}$ of compatibility structures, where $\{M_i\}_{i \in I}$ is a disjoint family of sets.

  Given such a structure, we form the disjoint union $M = \coprod_{i \in I} M_i$, and the compatibility structure $\mathcal{C}$ on M, where:

  $$\mathcal{C} = \{S \subseteq M \mid \forall i \in I.\, S_{(i)} \in \mathcal{C}_i\}.$$

  The idea is that measurements in different parts are always mutually compatible. Note that:

  $$S \subseteq S' \iff \forall i \in I.\, S_{(i)} \subseteq S'_{(i)},$$

  so $\mathbf{P}_{\mathcal{C}}$ is the product of the posets $\mathbf{P}_{\mathcal{C}_i}$.

Note that every compatibility structure $\mathcal{C}$ has bounded joins; in fact, it is closed under unions of bounded subsets. It is also closed under non-empty intersections.

3.2 Quantum Realizations of Contextual Posets

We shall now show how compatibility structures arise in the Quantum Mechanics formalism.

- Firstly, 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 M of observables, we form a compatibility structure by taking all finite subsets $S \subseteq M$ of pairwise commuting observables. Note that pairwise commutation implies that the observables in the subset, composed in any order, form a well-defined observable.

- Now suppose we have a list $\mathcal{H}_1, \ldots, \mathcal{H}_k$ of Hilbert spaces, and a set M_i of observables on $\mathcal{H}_i$, $i = 1, \ldots, k$. We make the observables in M_i 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 $A \in M_i$. Then A^i commutes with B^j whenever $i \neq j$, and we can form a compatibility structure $\mathcal{C}_{\mathsf{L}}$.
- Finally, we can combine these ideas in the natural fashion, using the commuting observables in each M_i to define the compatibility structure $\mathcal{C}_i$, and then combining these as local observables to form the compatibility structure $\mathcal{C}$.

3.3 Events

We shall assume given a sheaf $\mathcal{E} : \mathbf{P}^{\mathsf{op}} \to \mathbf{Set}$, which we think of as a sheaf of (atomic) **events**.

Example Each of the posets $\mathbf{P}$ from the examples given in 3.1 has a concrete representation as a family of sets of measurements. In each of these cases, we shall use the following event sheaf.

- We fix a set O of possible **outcomes** for each measurement. For each $S \in \mathbf{P}$, an **S-section** is a function $s : S \to O$. We interpret such an S-section as a specification of **possible outcomes** when the joint measurement S is performed. It should be read as a description of a history or run of the system, in which the measurements in the compatible family S were performed, and the outcomes $s(m)$ for each $m \in S$ were observed.

  The assignment $S \mapsto O^S$ on $\mathbf{P}$ becomes a functor $\mathcal{E}_O : \mathbf{P}^{\mathsf{op}} \to \mathbf{Set}$ as follows: given $S \subseteq S' \in \mathcal{C}$, an S'-section $s : S' \to O$ is mapped to the S-section $s|S : S \to O$.

  The presheaf $\mathcal{E}_O$ is in fact a sheaf, since under the compatibility conditions, we can glue the sections $s_j : S_j \to O$ together to form a section $s : S \to O$ on $S = \bigcup_j S_j$.

3.4 The Quantum Case

In the case where the measurements are observables on a Hilbert space, we shall assume that they each have the same standard set of possible outcomes. 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 $\mathsf{Tr}(\rho \mathbf{P}_i)$, 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 presheaf $\mathcal{E}_O$ to record the collective outcomes of a compatible set of observables.

Thus for each basic measurement m in M, 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 set of commuting observables $S = \{A_{m_1}, \ldots, A_{m_k}\}$, for each $s \in O^S$ we have a projector $\mathbf{P}_s = \mathbf{P}_{m_1}^{s(m_1)} \circ \cdots \circ \mathbf{P}_{m_k}^{s(m_k)}$. The composed observable $A_S = A_{m_1} \circ \cdots \circ A_{m_k}$ has a decomposition of the form

$$A_S = \sum_{s \in O^S} \alpha_s \mathbf{P}_s,$$

where $\alpha_s = \prod_i \alpha_{m_i}^{s(m_i)}$. 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:

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

4 Models

We now have the ingredients to set up a very general framework for studying structural properties of theories which make predictions about the outcomes of measurements.

We shall assume given:

- A poset $\mathbf{P}$ of contexts, which we assume to be bounded complete.
- A sheaf $\mathcal{E}$ of events on $\mathbf{P}$.
- A commutative, affine monad (T, η, μ, κ) which describes a notion of **effect** or **weight** which the theory ascribes to the possible events. Of particular interest will be the distribution monads, especially the probability monad $\mathcal{D}_{\mathsf{R}}$, and the boolean monad $\mathcal{D}_{\mathsf{B}}$.

With this language, we can formulate a very general notion of **empirical model**, which describes a scenario in which events can arise in various contexts; the model specifies the weights to be ascribed to the events. We will also be able, at the same level of generality, to formulate a notion of **hidden-variable model**, which makes use of some additional information encoded in a set of values for a hidden variable; and the crucial notion of how a hidden-variable model **realizes** an empirical model.

4.1 Empirical Models

An **empirical model** e assigns a weight on events: for each context $p \in \mathbf{P}$, this is an element e_p of $T(\mathcal{E})(p)$. We furthermore require that this assignment satisfies the following naturality property: whenever $p \leq p'$, $e_p = e_{p'}|p$. Here we extend the restriction notation to the presheaf $T(\mathcal{E})$: if $d \in T(\mathcal{E})(p')$ and $p \leq p'$,

$$d|p := T(\mathcal{E}(\iota_{p,p'}))(d).$$

This says exactly that e is a **global section** of the presheaf $T(\mathcal{E})$, *i.e.* a natural transformation $e : \mathbf{1} \overset{\cdot}{\longrightarrow} T(\mathcal{E})$, where we write $\mathbf{1}$ for the constant presheaf on $\mathbf{P}$ which assigns a fixed one-point set $\{\bullet\}$ (*i.e.* a terminal object in **Set**) to every $p \in \mathbf{P}$.

In the case of a distribution monad $\mathcal{D}_R$, we can write this restriction explicitly, as follows. For each $s \in \mathcal{E}(p)$:

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

Example When $\mathbf{P} = \mathbf{P}_{\mathcal{C}_{\mathsf{L}}}$, $\mathcal{E}_O$ is the presheaf of O-valued functions, and $T = \mathcal{D}_R$, this yields the following condition. Whenever $m_i \in S$, $i \in I$, $s \in \mathcal{E}(\{m_i\})$:

$$e_{\{m_i\}}(s) \;=\; \sum_{s' \in \mathcal{E}_O(S), s'|\{m_i\} = s} e_S(s').$$

This says that the distribution on a measurement outcome at part i is the **marginal** of the distribution on those measurement outcomes in the context S which restrict to this outcome at i. This is the **no-signalling** condition [16, 24], which thus emerges as a basic structural property of our framework.

4.2 The Quantum Case

We now consider the situation described in Section 3.4, where the measurements are quantum observables on a Hilbert space $\mathcal{H}$. A state ρ on $\mathcal{H}$ gives rise to an empirical model. For each compatible set of observables S, the state defines a probability distribution ρ_S over $\mathcal{E}_O(S)$, by the standard 'statistical algorithm' of quantum mechanics: $\rho_S(s) = \mathsf{Tr}(\rho \mathbf{P}_s)$. Thus $\rho_S \in \mathcal{D}_{\mathsf{R}}(\mathcal{E})(S)$ for each $S \in \mathcal{C}$. We must verify the naturality condition. If we are given $S_1 \subseteq S$, we can write $S = S_1 \cup S_2$, where $S_2 := S \setminus S_1$. Note that $O^S \cong O^{S_1} \times O^{S_2}$, so we can write $s \in \mathcal{E}_O(S)$ as

$s = (s_1, s_2)$, where $s_i \in O^{S_i}$, $i = 1, 2$. Moreover, restriction is just projection: $(s_1, s_2)|S_1 = s_1$. Now we can calculate:

$$\begin{array}{rcl} \rho_S|S_1(s_1) & = & \sum_{s_2 \in \mathcal{E}_O(S_2)} \rho_{S_1 \cup S_2}(s_1, s_2) \\ & = & \sum_{s_2 \in \mathcal{E}_O(S_2)} \mathsf{Tr}(\rho \mathbf{P}_{(s_1, s_2)}) \\ & = & \sum_{s_2 \in \mathcal{E}_O(S_2)} \mathsf{Tr}(\rho \mathbf{P}_{s_1} \mathbf{P}_{s_2}) \\ & = & \mathsf{Tr}(\sum_{s_2 \in \mathcal{E}_O(S_2)} \rho \mathbf{P}_{s_1} \mathbf{P}_{s_2}) \\ & = & \mathsf{Tr}(\rho \mathbf{P}_{s_1} \sum_{s_2 \in \mathcal{E}_O(S_2)} \mathbf{P}_{s_2}) \\ & = & \mathsf{Tr}(\rho \mathbf{P}_{s_1} I) \\ & = & \mathsf{Tr}(\rho \mathbf{P}_{s_1}) \\ & = & \rho_{S_1}(s_1). \end{array}$$

Although this derivation is straightforward, it makes an interesting point. The no-signalling condition extends to a condition relating to compatible families of observables in general, not just those represented as operating on different factors of a tensor product, and hence considered as possibly space-like separated. Indeed, from this perspective, we may say that the structural significance of the tensor product representation is precisely that it guarantees that local observables in different factors commute, with compatibility as the more fundamental phenomenon.

It should also be noted that this naturality property in no way precludes contextuality phenomena; this will become clear in what follows.

4.3 Hidden-Variable Models

We fix a set Λ of values for a hidden variable. A hidden-variable model h over Λ assigns, for each context $p \in \mathbf{P}$ and value $\lambda \in \Lambda$, a weight $h_p^\lambda \in T(\mathcal{E})(p)$. It also assigns a distribution $h_\Lambda \in T(\Lambda)$ on the hidden variables. The standard structural assumption of **λ-independence** (the term is from [12]) says that h_Λ is independent of the context. We also consider a naturality requirement corresponding to the one we imposed on empirical models: whenever $p \leq p'$, then $h_p^\lambda = h_{p'}^\lambda|p$.

These requirements say exactly that h is a natural transformation

$$h : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})^{\mathbf{\Lambda}} \times T(\Lambda)$$

where $\mathbf{\Lambda}$ is the constant presheaf valued at Λ. Thus λ-independence is encapsulated structurally by taking $\mathbf{\Lambda}$ to be a constant presheaf.

More precisely, the naturality condition for h, together with the fact that $\mathbf{\Lambda}$ is a constant presheaf, imply that $(h_\Lambda)_p = (h_\Lambda)_{p'}$ whenever $p \leq p'$. Since we are assuming that $\mathbf{P}$ is bounded complete, and in particular has a bottom element 0, and is therefore **connected**, then this implies that $(h_\Lambda)_p = (h_\Lambda)_0 = h_\Lambda$ for all $p \in \mathbf{P}$. Note that all compatibility structures as described in Section 3.1 have bottom elements, given by the empty set.

Example When $\mathbf{P} = \mathbf{P}_{\mathcal{C}_\mathsf{L}}$, $\mathcal{E}_O$ is the presheaf of O-valued functions, and $T = \mathcal{D}_R$, the naturality requirement on h yields the following condition. Whenever $m_i \in S$, $i \in I$, $s \in \mathcal{E}(\{m_i\})$:

$$h_{\{m_i\}}^\lambda(s) \;=\; \sum_{s' \in \mathcal{E}_O(S), s'|\{m_i\} = s} h_S^\lambda(s').$$

This says that, for each fixed value of the hidden variable, the distribution on measurement outcomes at part i is the marginal of the distribution on those measurement outcomes in the context S which restrict to this outcome at i. This is the **parameter independence** condition [40, 23]. Thus, λ-independence and parameter independence emerge as basic structural properties of hidden-variable models in this framework.

4.4 Realization and Equivalence

We shall now show how the notion of a hidden-variable model **realizing** an empirical model arises canonically in this setting. What we want is a morphism of presheaves

$$\mathsf{kev}^T : T(\mathcal{E})^{\mathbf{\Lambda}} \times T(\Lambda) \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$$

since then we can simply define $e := \mathsf{kev}^T \circ h : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$. However, the required morphism kev^T is simply the **Kleisli evaluation** which arises automatically from the commutative monad structure of T and the cartesian closed structure of $\mathbf{Set}^{\mathbf{P}^{\mathsf{op}}}$.

Indeed, given $t : \mathbf{\Lambda} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$, the **Kleisli extension** of t is the morphism

$$t^\sharp : T(\mathbf{\Lambda}) \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$$

defined by

$$t^\sharp = T(\mathbf{\Lambda}) \xrightarrow{T(t)} T(T(\mathcal{E})) \xrightarrow{\mu_{\mathcal{E}}} T(\mathcal{E}).$$

In the case of the distribution monads $\mathcal{D}_R$, this is defined explicitly as follows. Firstly, for each $\lambda \in \Lambda$ and $p \in \mathbf{P}$, t assigns a distribution $t_p^\lambda \in \mathcal{D}_R(\mathcal{E})(p)$. Now

$$t_p^\sharp :: d_\Lambda \mapsto [s \mapsto \sum_{\lambda\in\Lambda} d_\Lambda(\lambda) \cdot t_p^\lambda(s)].$$

The Kleisli evaluation kev^T is the internalized version of this process obtained using the cartesian closed structure of $\mathbf{Set}^{\mathbf{P}^{\mathsf{op}}}$. For a general description, suppose (T, η, μ, κ) is a commutative monad on a cartesian closed category $\mathcal{C}$. We can define the Kleisli evaluation as follows:

$$\mathsf{kev}^T_{X,Y} = TX^Y \times TY \xrightarrow{\eta_{TX^Y}\times 1} T(TX^Y) \times TY \xrightarrow{\kappa_{TX^Y,Y}} T(TX^Y \times Y) \xrightarrow{T\mathsf{ev}_{TX,Y}} TTX \xrightarrow{\mu_X} TX.$$

In our context, we can take advantage of the simplified form which this Kleisli evaluation takes, since the exponent $\mathbf{\Lambda}$ is a constant presheaf. In the case of the distribution monads $\mathcal{D}_R$, it is given explicitly by

$$\mathsf{kev}_p^{\mathcal{D}_R} : \mathcal{D}_R(\mathcal{E})^{\mathbf{\Lambda}}(p) \times \mathcal{D}_R(\Lambda) \longrightarrow \mathcal{D}_R(\mathcal{E})(p) :: ((h_p^\lambda)_{\lambda\in\Lambda}, h_\Lambda) \mapsto [s \mapsto \sum_{\lambda\in\Lambda} h_\Lambda(\lambda) \cdot h_p^\lambda(s)].$$

This can be recognized as the standard expression for how observable behaviour is realized by averaging over the values of the hidden variable.

We say that two hidden-variable models h, h', over the same monad T, but quite possibly with different sets of hidden variables, are **equivalent** if they realize the same empirical model; that is, if $\mathsf{kev}^T \circ h = \mathsf{kev}^T \circ h'$.

5 Properties of Models

Some of the key properties of empirical and hidden-variable models which have been considered in the literature, notably no-signalling, parameter independence and λ-independence, have been shown to arise directly from basic structural conditions in the sheaf formalism. We can think of them as **framework properties** in our setting.

We shall now formulate two key properties in the sheaf-theoretic language. The first, which we call factorizability, can be seen as a common generalization of locality and non-contextuality. The second is determinism. These both arise in a natural and compelling fashion, and can in fact be expressed in great generality in the sheaf framework.

5.1 Factorizability

We shall formulate a general independence property, and show that it yields locality and non-contextuality as standardly considered in quantum foundations, when we specialize it appropriately.

Suppose we have a finite disjoint cover $\{p_j\}_{j\in J}$ of $p \in \mathbf{P}$. By Proposition 2.2, $\mathcal{E}(p) \cong \prod_{j\in J} \mathcal{E}(p_j)$. From the commutativity of the monad, we have a morphism

$$\kappa_p^{\mathbf{\Lambda}} : \prod_{j\in J} T(\mathcal{E})^{\mathbf{\Lambda}}(p_j) \longrightarrow T(\prod_{j\in J} \mathcal{E}(p_j))^{\Lambda} \;\cong\; T(\mathcal{E})^{\mathbf{\Lambda}}(p).$$

We say that a hidden-variable model $h : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})^{\mathbf{\Lambda}} \times T(\mathbf{\Lambda})$ satisfies **factorizability** if it factors through this morphism, as

$$h_p \;=\; \kappa_p^{\mathbf{\Lambda}} \circ \langle\, h_{p_j} \mid j \in J\rangle. \tag{2}$$

If we apply this general definition in the case of the distribution monads, we obtain the following condition. Whenever $\{p_j\}_{j\in J}$ is a finite disjoint cover of p, then for all $\lambda \in \Lambda$ and $s \in \mathcal{E}(p)$:

$$h_p^{\lambda}(s) \;=\; \prod_{j\in J} h_{p_j}^{\lambda}(s{\restriction}p_j). \tag{3}$$

Note that in this expression, the product is that of the semiring!

The hypothesis of finiteness of the disjoint cover is needed in order for the product in (3) to be well-defined. More generally, the commutative action κ is only defined for finite products. We shall discuss the possibilities for generalization in the Concluding Remarks. In our concrete examples of compatibility structures, the contexts will generally be finite subsets of the underlying set of basic measurements, so finiteness of disjoint covers will hold automatically.

Examples Suppose that $\mathbf{P} = (\mathcal{C}, \subseteq)$ for some compatibility structure $\mathcal{C}$, $\mathcal{E} = \mathcal{E}_O$, and $T = \mathcal{D}_R$.

- If $\mathcal{C} = \mathcal{C}_{\mathsf{L}}$, then the condition specializes to **Bell locality** [4]. That is, for each value of the hidden variable, the distribution on the outcomes factors as a product of distributions on each site or local part of the system.
- For compatibility structures in general, the condition yields a form of **non-contextuality**. This is less standard, and so we shall spell it out in more detail.

We say that a hidden-variable model h is **decomposable** if for each $\lambda \in \Lambda$ there is a distribution $d_\lambda \in \mathcal{D}_R(M \times O)$ such that, for all $S \in \mathcal{C}$:

$$h_S^{\lambda}(s) = \prod_{m\in S} d_\lambda(m, s(m)).$$

Note that this condition says that the joint distribution factors into independent distributions determined 'locally' (without any necessary spatial connotation) for each basic measurement, independent of the context in which it appears.

The following result is a special case of the Extension Theorem 6.3 which we will prove in Section 6.

Proposition 5.1 *When T is a distribution monad, a hidden-variable model h is factorizable if and only if it is decomposable.*

We shall say that an empirical model which has no realization by a hidden-variable model satisfying factorizability is **contextual**.

Outcome Independence

In the literature on hidden-variable models, locality is factored into parameter independence and outcome independence [23, 40]. As we have already seen, parameter independence is a basic requirement for hidden-variable models in our setting — it is the naturality requirement for the hidden-variable model to be a section of the presheaf $\mathcal{D}_R(\mathcal{E})^{\mathbf{\Lambda}} \times \mathcal{D}_R(\Lambda)$.

We can formulate **outcome independence** for distribution monads as follows. Whenever $\{p_j\}_{j \in J}$ is a disjoint cover of p, then for all $\lambda \in \Lambda$ and $s \in \mathcal{E}(p)$:

$$h_p^{\lambda}(s) = \prod_{j \in J} \sum_{\{s' \in \mathcal{E}(p) | s' \upharpoonright p_j = s \upharpoonright p_j\}} h_p^{\lambda}(s').$$

This says that the outcomes in a given context are independent of each other. The following is now immediate.

Proposition 5.2 *For each hidden-variable model $h : \mathbf{1} \stackrel{.}{\longrightarrow} \mathcal{D}_R(\mathcal{E})^{\mathbf{\Lambda}} \times \mathcal{D}_R(\Lambda)$, which by naturality satisfies parameter independence, h satisfies factorizability if and only if it additionally satisfies outcome independence.*

5.2 Determinism

A hidden-variable model h is **deterministic** if it factors as

$$h \ = \ \mathbf{1} \xrightarrow{\langle x, h_\Lambda \rangle} \mathcal{E}^{\mathbf{\Lambda}} \times T(\Lambda) \xrightarrow{\eta^{\mathbf{\Lambda}} \times 1} T(\mathcal{E})^{\mathbf{\Lambda}} \times T(\Lambda). \tag{4}$$

For distribution monads, this says that for each $\lambda \in \Lambda$, there is $x^{\lambda} : \mathbf{1} \stackrel{.}{\longrightarrow} \mathcal{E}$ such that for all p, $h_p^{\lambda} = \delta_{x_p^{\lambda}}$.

5.3 Summary of Definitions

It may be helpful to summarize the main definitions at this stage. The ingredients of our approach in full generality are:

- A bounded complete poset $\mathbf{P}$ of contexts.
- A sheaf $\mathcal{E}$ over $\mathbf{P}$.
- An affine commutative monad (T, η, μ, κ) on $\mathbf{Set}$, which lifts to an affine commutative monad on $\mathbf{Set}^{\mathbf{P}^{\mathsf{op}}}$.

In these terms:

- An empirical model is a natural transformation $e : \mathbf{1} \stackrel{.}{\longrightarrow} T(\mathcal{E})$.
- A hidden-variable model is a natural transformation $h : \mathbf{1} \stackrel{.}{\longrightarrow} T(\mathcal{E})^{\mathbf{\Lambda}} \times T(\mathbf{\Lambda})$, where Λ is a constant presheaf.
- A hidden-variable model h realizes an empirical model e if $e = \mathsf{kev}^{T}_{\mathcal{E},\mathbf{\Lambda}} \circ h$.
- A hidden-variable model is factorizable if (2) holds for every disjoint cover.
- A hidden-variable model is deterministic if (4) holds.

5.4 Determinism Implies Factorizability

We shall now show, for any poset $\mathbf{P}$, and any commutative monad (T, η, μ, κ), that if a hidden-variable model $h : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})^{\mathbf{\Lambda}} \times T(\mathbf{\Lambda})$ is deterministic, then it is factorizable.

Proposition 5.3 *If $h : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})^{\mathbf{\Lambda}} \times T(\mathbf{\Lambda})$ is deterministic, it is factorizable.*

Proof Recall that h deterministic means that for each $\lambda \in \Lambda$, there is $x^{\lambda} : \mathbf{1} \stackrel{\cdot}{\longrightarrow} \mathcal{E}$ such that $h^{\lambda}_{p} = \eta_{p} \circ x^{\lambda}_{p}$. Now let $\{p_j\}$ be a disjoint cover of p. Since $\mathcal{E}$ is a sheaf, $\mathcal{E}(p) \cong \prod_{j \in J} \mathcal{E}(p_j)$ by Proposition 2.2. By coherence of κ:

$$\begin{array}{rcl} h^{\lambda}_{p} & = & \eta_{\prod_j \mathcal{E}(p_j)} \circ \langle x^{\lambda}_{p_j} \mid j \in J \rangle \\ & = & \kappa_{\mathcal{E}(p)} \circ \prod_j \eta_{\mathcal{E}(p_j)} \circ \langle x^{\lambda}_{p_j} \mid j \in J \rangle \\ & = & \kappa_{\mathcal{E}(p)} \circ \langle \eta_{\mathcal{E}(p_j)} \circ x^{\lambda}_{p_j} \mid j \in J \rangle \\ & = & \kappa_{\mathcal{E}(p)} \circ \langle h^{\lambda}_{p_j} \mid j \in J \rangle . \end{array}$$

□

6 Structure of Factorizability

In this section, we analyze the mathematical structure of Factorizability. We shall begin by considering some additional requirements on our poset of contexts $\mathbf{P}$.

An **atom** in a poset $\mathbf{P}$ with least element 0 is an element $a \in P$, $a \neq 0$, such that, for all $p \in \mathbf{P}$, if $p \leq a$ then $p = 0$ or $p = a$. We write $\mathbf{A}$ for the set of atoms in $\mathbf{P}$.

A finite poset $\mathbf{P}$ is **simplicial** [6] if it has a bottom element 0, and for every $p \in \mathbf{P}$, the interval $[0, p]$ is a Boolean algebra. Note that a finite Boolean algebra B can be represented as the powerset of its atoms: $B \cong \mathcal{P}(\mathbf{A})$.

Examples The posets arising from finite compatibility structures $\mathcal{C}$ are simplicial. Note that $\mathcal{C}$ is finite if and only if the underlying set M is finite.

In fact, we can show that finite compatibility structures are fully representative of simplicial posets.

Proposition 6.1 (Representation Theorem) *A poset is simplicial if and only if it is isomorphic to a finite compatibility structure.*

Proof Clearly, finite compatibility structures are simplicial. For the converse, suppose that $\mathbf{P}$ is simplicial, with set of atoms $\mathbf{A}$. We write $\mathbf{A}(p)$ for the set of atoms below $p \in \mathbf{P}$. We form the compatibility structure $\mathcal{C}$ on $\mathbf{A}$ comprising the bounded sets of atoms. $\mathcal{C}$ is finite, since $\mathbf{P}$ and hence $\mathbf{A}$ is, and it is evidently downwards closed. We define a map

$$\phi : \mathbf{P} \rightarrow \mathcal{C} :: p \mapsto \mathbf{A}(p).$$

Since for all $p \in \mathbf{P}$, $p = \bigvee \mathbf{A}(p)$, this map is an order-embedding:

$$p \leq p' \iff \phi(p) \subseteq \phi(p').$$

Finally, since the downset of each element of $\mathbf{P}$ is a finite Boolean algebra, it has a unique representation as the join of a set of atoms. Hence if $\bigvee S = \bigvee S'$, we must have $S = S'$. Thus $S = \phi(\bigvee S)$, and ϕ is surjective. □

We also note the following evident special case.

Proposition 6.2 *If a simplicial poset has a top element, it is a finite Boolean algebra.*

For the remainder of this Section, we shall assume that $\mathbf{P}$ is simplicial. We write $\mathbf{A}$ for the set of atoms of $\mathbf{P}$, and $\mathbf{A}(p)$ for the set of atoms below $p \in \mathbf{P}$. Note that $\mathbf{A}(p)$ forms a disjoint cover of p.

Suppose we are given a family $g = \{g^\lambda_a\}$, where $g^\lambda_a \in T(\mathcal{E})(a)$, $a \in \mathbf{A}$, $\lambda \in \Lambda$. We can extend this family to an assignment $g^\lambda_p \in T(\mathcal{E})(p)$, $p \in \mathbf{P}$, by

$$(g^\lambda_p)_{\lambda\in\Lambda} = \kappa^{\mathbf{\Lambda}}_p \circ \langle (g^\lambda_a)_{\lambda\in\Lambda} \mid a \in \mathbf{A}(p)\rangle.$$

We shall now show that this extension of the assignment to atoms satisfies the naturality requirement to be a model. Moreover, all factorizable models arise in this way.

Proposition 6.3 (Extension Theorem) *Assume that $\mathbf{P}$ is simplicial. An assignment on atoms $g = \{g^\lambda_a\}$, where $g^\lambda_a \in T(\mathcal{E})(a)$, $a \in \mathbf{A}$, $\lambda \in \Lambda$, extends to a morphism $g : \mathbf{1} \overset{\cdot}{\longrightarrow} T(\mathcal{E})^{\mathbf{\Lambda}}$ by*

$$(g^\lambda_p)_{\lambda\in\Lambda} = \kappa^{\mathbf{\Lambda}}_p \circ \langle (g^\lambda_a)_{\lambda\in\Lambda} \mid a \in \mathbf{A}(p)\rangle. \tag{5}$$

This morphism satisfies factorizability. Conversely, a hidden-variable model h which satisfies factorizability is the extension by (5) of its restriction to atoms.

Proof We shall take advantage of the Representation Theorem 6.1 and write elements of the poset as sets of atoms. Thus an order relation in the poset has the form $S \subseteq S \cup S'$, where $S \cap S' = \varnothing$. By Proposition 2.2, $\mathcal{E}(S \cup S') \cong \mathcal{E}(S) \times \mathcal{E}(S')$. Moreover, we have

$$\mathcal{E}(\iota_{S,S\cup S'}) \;=\; \mathcal{E}(S) \times \mathcal{E}(S') \xrightarrow{\pi_1} \mathcal{E}(S), \tag{6}$$

i.e. the restriction map $\mathcal{E}(\iota_{S,S\cup S'})$ is the projection $\pi_1 : \mathcal{E}(S) \times \mathcal{E}(S') \to \mathcal{E}(S)$. Indeed, since the members of S form a disjoint cover of it, with the restrictions to atoms as projections, we have, for each $a \in S$,

$$\mathcal{E}(\iota_{\{a\},S\cup S'}) \;=\; \prod_{a' \in S\cup S'} \mathcal{E}(a') \xrightarrow{\pi_a} \mathcal{E}(a),$$

and by functoriality of $\mathcal{E}$ we have $(s{\restriction}S){\restriction}a = s{\restriction}a$, so by the universal property of the product we must have $s{\restriction}S = (s{\restriction}a \mid a \in S)$.

Now suppose we are given an assignment to atoms $g = \{g^\lambda_a\}$, and $S \subseteq S \cup S'$. We must show that $g^\lambda_{S\cup S'}{\restriction}S = g^\lambda_S$. This is now a calculation:

$$\begin{array}{rcll}
g^\lambda_{S\cup S'}{\restriction}S & = & T(\mathcal{E}(\iota_{S,S\cup S'})) \circ \kappa_{S\cup S'}(g^\lambda_a \mid a \in S \cup S') & \\
& = & T(\pi_1) \circ \kappa_{S\cup S'}(g^\lambda_a \mid a \in S \cup S') & (6) \\
& = & T(\pi_1) \circ \kappa_{S,S'}(\kappa_S(g^\lambda_a \mid a \in S), \kappa_{S'}(g^\lambda_a \mid a \in S')) & \text{Coherence of } \kappa \\
& = & \pi_1(\kappa_S(g^\lambda_a \mid a \in S), \kappa_{S'}(g^\lambda_a \mid a \in S')) & T \text{ is affine} \\
& = & \kappa_S(g^\lambda_a \mid a \in S) & \\
& = & g^\lambda_S. &
\end{array}$$

To show factorizability, suppose we have a disjoint cover $\{S_1, \ldots, S_n\}$ of S; that is, a disjoint family with $\bigcup S_i = S$. Then, using the coherence of κ:

$$\begin{array}{rcl}
g^\lambda_S & = & \kappa_S(g^\lambda_a \mid a \in S) \\
& = & \kappa_{S_1\ldots,S_n}(\kappa_{S_1}(g^\lambda_a \mid a \in S_1), \ldots, \kappa_{S_n}(g^\lambda_a \mid a \in S_n)) \\
& = & \kappa_{S_1\ldots,S_n}(g^\lambda_{S_1}, \ldots, g^\lambda_{S_n}).
\end{array}$$

For the converse, we note that (5) is an instance of factorizability, with respect to the disjoint cover $\mathbf{A}(p)$ of p. $\square$

We define a subset $\mathcal{L}(p) \subseteq T(\mathcal{E})^{\mathbf{\Lambda}}(p)$ for each $p \in \mathbf{P}$, as the set of all g_p^λ arising from some assignment g on atoms by (5).

Proposition 6.4 *The family $\{\mathcal{L}(p)\}$ determines a sub-presheaf of $T(\mathcal{E})^{\mathbf{\Lambda}}$. This presheaf is a sheaf.*

Proof Given $g_{p'}^\lambda \in \mathcal{L}(p')$ and $p \leq p'$, $g_{p'}^\lambda \restriction p = g_p^\lambda \in \mathcal{L}(p)$ by Proposition 6.3. This shows the naturality of the inclusions $\mathcal{L}(p) \hookrightarrow T(\mathcal{E})^{\mathbf{\Lambda}}(p)$, and hence that $\mathcal{L}$ is a sub-presheaf.

Now suppose we have a covering $\{p_j\}$ of p, and a compatible family $s_j \in \mathcal{L}(p_j)$. Each s_j can be written as

$$s_j^\lambda = \kappa_{p_j}(g_a^\lambda \mid a \in \mathbf{A}(p_j)).$$

The compatibility of the s_j means that we obtain a single consistent assignment $\{g_a^\lambda \mid a \in \mathbf{A}(p)\}$. This then determines $(\kappa_p(g_a^\lambda \mid a \in \mathbf{A}(p)))_{\lambda \in \Lambda} \in \mathcal{L}(p)$, and this assignment is unique by Proposition 6.3. $\square$

We refer to $\mathcal{L}$ as the **sheaf of local models**. The following is an immediate consequence of Propositions 6.3 and 6.4.

Proposition 6.5 *If $\mathbf{P}$ is simplicial, a hidden-variable model satisfies factorizability if and only if it is a global section of the sheaf of local models; that is, if and only if it factors through the inclusion $\mathcal{L} \hookrightarrow T(\mathcal{E})^{\mathbf{\Lambda}}$.*

7 Determinization

We shall now show that, if there are no constraints on the hidden variables, then it is always possible to introduce deterministic hidden variables for any empirical model, defined on an arbitrary commutative monad in any cartesian closed category.

We say that a morphism $p : A \to B$ in a category $\mathcal{C}$ has the **lifting property** if every $f : C \to B$ factors through p; *i.e.* for some $g : C \to A$, (not required to be unique), $f = p \circ g$. Diagrammatically:

$$\begin{tikzcd} & A \arrow[d, "p"] \\ C \arrow[ur, dashed, "g"] \arrow[r, "f"'] & B \end{tikzcd}$$

Given an arrow $f : C \times A \to B$, we write $\lambda(f) : C \to B^A$ for its exponential transpose. We also have the **strength map** $\mathsf{st}_{A,B} : B^A \to TB^{TA}$, defined as the exponential transpose of:

$$B^A \times TA \xrightarrow{\eta \times 1} T(B^A) \times TA \xrightarrow{\kappa} T(B^A \times A) \xrightarrow{T\mathsf{ev}} TB.$$

Note that the following diagram commutes:

$$\begin{tikzcd} X^Y \times TY \arrow[r, "\mathsf{st}_{X,Y} \times 1"] \arrow[d, "\eta^Y \times 1"'] & TX^{TY} \times TY \arrow[d, "\mathsf{ev}"] \\ TX^Y \times TY \arrow[r, "\mathsf{kev}"'] & TX \end{tikzcd}$$

so we can use the strength to write the Kleisli evaluation in a simplified form in the case of a deterministic model.

Now let T be a commutative monad on a cartesian closed category $\mathcal{C}$. We say that T admits **deterministic hidden variables** if for every object A there is an object $\mathcal{V}$ such that the arrow

$$\mathsf{ev} \circ (\mathsf{st} \times 1) : A^{\mathcal{V}} \times T(\mathcal{V}) \to T(A)$$

has the lifting property. Note that this says in particular that, for every $e : \mathbf{1} \to T(A)$ — every empirical model — there exists $h : \mathbf{1} \to A^{\mathcal{V}} \times T(\mathcal{V})$ — a hidden-variable model — such that $\mathsf{ev} \circ (\mathsf{st} \times 1) \circ h = e$ — *i.e.* such that h realizes e.

Proposition 7.1 (Determinization Theorem) *Every commutative monad on a cartesian closed category $\mathcal{C}$ admits deterministic hidden variables.*

Proof We take $\mathcal{V} := A$. Given $f : C \to TA$, we define h by

$$h := \langle \lambda(\pi_2), f \rangle.$$

Now $\mathsf{st}_{A,A}(\lambda(\pi_2^{B,B})) = \lambda(\pi_2^{TB,TB})$, so

$$\mathsf{ev} \circ (\mathsf{st} \times 1) \circ h = \mathsf{ev} \circ (\mathsf{st} \times 1) \circ \langle \lambda(\pi_2), f \rangle = \mathsf{ev} \circ (\lambda(\pi_2) \times 1) \circ \langle 1, f \rangle = \pi_2 \circ \langle 1, f \rangle = f.$$

Thus h lifts f. $\square$

Note the tautological nature of this construction: we use the set of 'events' as its own hidden variable set. Since $\mathbf{Set}^{\mathbf{P}^{\mathsf{op}}}$ is always cartesian closed, this might seem a rather striking result. What prevents us using it to obtain a deterministic hidden variable realization of every empirical model? One thing and one thing only: λ-independence! Namely, in presheaf terms, the constraint that the hidden variable object should be a **constant** presheaf. This then creates the necessary combination of varying contexts of measurements, and fixed values of hidden variables we can thread through these different contexts — a notable feature of the proofs of No-Go theorems.

Thus not only do we have the anecdotal experience that one always needs λ-independence to prove any No-Go results involving hidden variables; this result is telling us, in very general terms, that without a constraint such as λ-independence, we can always find — trivially — a deterministic hidden variable theory.

8 Extendability and Factorizability

We now consider the situation where are all measurements are compatible, and hence can be performed jointly. Abstractly, this corresponds to the existence of a greatest element $\top$ of $\mathbf{P}$.

We show that for empirical models, compatibility implies determinization, and hence, by Proposition 5.3, factorizability. The idea is to determinize at the top as in the Determinization Theorem 7.1, and then everything else is forced by naturality.

Proposition 8.1 *Let $e : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$ be an empirical model. If $\mathbf{P}$ has a greatest element, e is realized by a deterministic hidden-variable model d.*

Proof We set $\Lambda := \mathcal{E}(\top)$, and define, for each $s \in \mathcal{E}(\top)$, $p \in \mathbf{P}$, $x_p^s := s{\restriction}p$ and $d_p^s := \eta_{\mathcal{E}(p)} \circ x_p^s$. This yields a natural transformation $d : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})^{\mathbf{\Lambda}}$, which is deterministic by construction. Moreover, we define $d_\Lambda := e_\top : \mathbf{1} \stackrel{\cdot}{\longrightarrow} \mathbf{\Lambda}$. It remains to verify that d realizes e. Since d is deterministic,

$$\mathsf{kev}^T \circ \langle d, d_\Lambda \rangle = \mathsf{ev} \circ (\mathsf{st} \times 1) \circ \langle x, d_\Lambda \rangle.$$

Now for any $p \in \mathbf{P}$, consider the following diagram:

$$\begin{tikzcd}
\mathbf{1} \arrow[r, "\langle x_\top, d_\Lambda \rangle"] \arrow[dr, "\langle x_p, d_\Lambda \rangle"'] & \mathcal{E}(\top)^{\mathcal{E}(\top)} \times T\mathcal{E}(\top) \arrow[r, "\mathsf{st} \times 1"] \arrow[d, "\mathcal{E}(\iota_{p,\top})^{\mathcal{E}(\top)} \times 1"] & T\mathcal{E}(\top)^{T\mathcal{E}(\top)} \times T\mathcal{E}(\top) \arrow[r, "\mathsf{ev}"] \arrow[d, "T\mathcal{E}(\iota_{p,\top})^{T\mathcal{E}(\top)} \times 1"] & T\mathcal{E}(\top) \arrow[d, "T\mathcal{E}(\iota_{p,\top})"] \\
& \mathcal{E}(p)^{\mathcal{E}(\top)} \times T\mathcal{E}(\top) \arrow[r, "\mathsf{st} \times 1"'] & T\mathcal{E}(p)^{T\mathcal{E}(\top)} \times T\mathcal{E}(\top) \arrow[r, "\mathsf{ev}"'] & T\mathcal{E}(p)
\end{tikzcd}$$

The squares commute because of the naturality of st and ev with fixed exponents; the triangle commutes because of the naturality of x. Now $x_\top = \lambda(1_{\mathcal{E}(\top)})$, and $\mathsf{st} \circ x_\top = \lambda(1_{T\mathcal{E}(\top)})$, so $\mathsf{ev} \circ (\mathsf{st} \times 1) \circ \langle x_\top, d_\Lambda \rangle = d_\Lambda = e_\top$. Hence $\mathsf{ev} \circ (\mathsf{st} \times 1) \circ \langle x_p, d_\Lambda \rangle = e_\top | p = e_p$. $\square$

It may be useful to write the proof out again in a more concrete fashion, in the special case when T is a distribution monad.

Proposition 8.2 *Let $e : \mathbf{1} \stackrel{\cdot}{\longrightarrow} \mathcal{D}_R(\mathcal{E})$ be an empirical model. If $\mathbf{P}$ has a greatest element, e is realized by a deterministic hidden-variable model d.*

Proof We introduce a hidden variable λ_s for every s in the support of $e_\top$. Thus Λ is an isomorphic copy of the support of $e_\top$. We define $d_p^{\lambda_s} = \delta_{s|p}$, and $d_\Lambda(\lambda_s) = e_\top(s)$. Clearly these are well-defined distributions. When $p \leq p'$, $(\delta_{s|p'})|p = \delta_{s|p}$, by naturality of η and functoriality of $\mathcal{E}$. Hence $d : \mathbf{1} \stackrel{\cdot}{\longrightarrow} \mathcal{D}_R(\mathcal{E})^\Lambda \times \mathcal{D}_R(\Lambda)$ is well-defined, and deterministic by construction. Finally, d realizes e, since, for all $p \in \mathbf{P}$, $s \in \mathcal{E}(p)$:

$$\sum_{\lambda_{s'} \in \Lambda} d_\Lambda(\lambda_{s'}) \cdot d_p^{\lambda_{s'}}(s) \;=\; \sum_{\lambda_{s'} \in \Lambda} e_\top(s') \cdot d_p^{\lambda_{s'}}(s) \;=\; \sum_{\lambda_{s'} \in \Lambda, s'|p = s} e_\top(s') \cdot 1 \;=\; \sum_{s'|p=s} e_\top(s') \;=\; e_\top|p\,(s),$$

and by naturality (no signalling), $e_p = e_\top|p$. $\square$

Note that we have no hypotheses on $\mathbf{P}$ other than the existence of a greatest element.

We now consider a form of converse to this result. Under the assumption that $\mathbf{P}$ is simplicial, we can show that, if a model is factorizable, then it can be extended to a model on the poset consisting of **all** sets of measurements, *i.e.* in which all measurements are compatible.

Proposition 8.3 *Assume that $\mathbf{P}$ is simplicial, and suppose that $e : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$ is realized by a factorizable hidden-variable model. Then e extends to an empirical model over the boolean algebra $\mathbf{Q} = \mathcal{P}(\mathbf{A})$ consisting of all sets of atoms of $\mathbf{P}$.*

Proof Firstly, by Proposition 2.2, we can extend $\mathcal{E}$ to $\mathcal{E}'$ defined on $\mathbf{Q}$, uniquely up to unique natural isomorphism, by:

$$\mathcal{E}'(q) = \prod_{a \in \mathbf{A}(q)} \mathcal{E}(a), \qquad q \in \mathbf{Q}.$$

Now suppose that e is realized by a factorizable hidden-variable model h. By the Extension Theorem 6.3, h is uniquely determined by its restriction to atoms $\{g_a^\lambda\}$; moreover, we can apply (5) to obtain an assignment g_q^λ on all $q \in \mathbf{Q}$, which defines a morphism $g : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E}')^{\mathbf{\Lambda}}$, which is factorizable, and which extends h. The hidden-variable model (g, h_Λ) realizes an empirical model e', defined on $\mathbf{Q}$, which extends e. $\square$

Combining these results, we obtain the following theorem.

Theorem 8.4 (E = F Theorem) *Assume that* $\mathbf{P}$ *is simplicial. The following are equivalent for empirical models* e*:*

1. *There is a realization of* e *by a hidden-variable model satisfying factorizability.*

2. *There is an extension of* e *to all sets of measurements.*

Proof (1) $\Rightarrow$ (2) is immediate from Proposition 8.3.

Conversely, (2) implies that e extends to a model on $\mathcal{P}(M)$, which has a top element, namely M. Now Proposition 8.1 implies that this extended model e' has a realization by a deterministic hidden-variable model h'. If we restrict back to $\mathcal{C}$, we get a deterministic hidden-variable model h which realizes e. Thus (2) $\Rightarrow$ (1). $\square$

To put this more colloquially, under suitable hypotheses:

Factorizability is equivalent to Extendability.

It may also be worthwhile to state the contrapositive:

Contextuality is equivalent to the existence of incompatible measurements.

9 Determinization and Relational Collapse

In this Section, we shall use our general setting to develop some general results which will provide a basis for proving No-Go theorems for a wide class of models.

9.1 Determinization for Factorizable Models

Proofs of No-Go Theorems often make use, explicitly as in [7, 1], or tacitly as e.g. in [32, 33], of **determinization results**, which say that if an empirical model is realized by a hidden-variable model with certain properties — typically some form of locality — then it can be realized by a hidden-variable model with the same properties, which is moreover **deterministic**. Results of this kind clearly also have considerable conceptual interest, as they bear on the question of whether the observable phenomena could, even in principle, be explained by a deterministic theory of a certain form.

Proposition 9.1 (Determinization for Factorizable Models) *Assume that* $\mathbf{P}$ *is simplicial. For any factorizable hidden-variable model* h*, there is an equivalent deterministic hidden-variable model.*

Proof Let e be the empirical model realized by h. By Proposition 8.3, e extends to a model e' defined on all sets of atoms. Since $\mathbf{P}$ is finite, its set of atoms $\mathbf{A}$ is also, so e' is defined on the top element of the extended poset $\mathcal{P}(\mathbf{A})$. Hence by Proposition 8.1, e' is realized by a deterministic hidden-variable model d'. Restricting d' to $\mathbf{P}$, we obtain a deterministic hidden-variable model which realizes e, and is therefore equivalent to h. $\square$

9.2 Contextuality by Restriction and Descent

We will discuss two powerful methods for lifting No-Go theorems.

Firstly, we note that to show contextuality, it is sufficient to show it for a small part of a model. Suppose we have an empirical model $e : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$ defined on a bounded complete poset $\mathbf{P}$. Let $\mathbf{P}'$ be a sub-poset of $\mathbf{P}$. There is a restriction e' of e to this sub-poset.

We say that $\mathbf{P}'$ is a **rigid** sub-poset of $\mathbf{P}$ if it is closed under bounded joins and binary meets from $\mathbf{P}$:

- If $S \subseteq \mathbf{P}'$, and S is bounded in $\mathbf{P}$, then $\bigvee_{\mathbf{P}} S \in \mathbf{P}'$.

- If $p, p' \in \mathbf{P}'$, so is $p \wedge_{\mathbf{P}} p'$.

Proposition 9.2 (No-Go Restriction Theorem) *Let $\mathbf{P}'$ be a rigid sub-poset of $\mathbf{P}$, and e an empirical model defined on $\mathbf{P}$. If the restriction of e to $\mathbf{P}'$ is contextual, so is e.*

Proof Suppose for a contradiction that e has a realization by a factorizable hidden-variable model h. The restriction of h to $\mathbf{P}'$ is a hidden-variable model which realizes the restriction of e to $\mathbf{P}'$. Note that, by rigidity of $\mathbf{P}'$, any disjoint cover in $\mathbf{P}'$ is a disjoint cover in $\mathbf{P}$. Hence the factorizability of h implies that of its restriction to $\mathbf{P}'$, contradicting the assumption of contextuality for e'. $\square$

Thus to prove the contextuality of a model defined on a large poset $\mathbf{P}$, it suffices to prove it for a small one which can be embedded rigidly into $\mathbf{P}$. In particular, this allows our results proved under the assumption that the poset of contexts is simplicial to be leveraged to apply much more widely.

Example Consider a compatibility structure $\mathcal{C}$ over a set M. Given any subset $M_0 \subseteq M$, we have the induced compatibility structure $\mathcal{C}_0 := \{S \in \mathcal{C} \mid S \subseteq M_0\}$. Note that $\mathcal{C}_0$ is finite if and only if M_0 is. Also, $\mathcal{C}_0$ is closed under bounded unions and non-empty intersections, and hence is a rigid sub-poset of $\mathcal{C}$.

For example, we have the compatibility structure where M is the set of all observables on a Hilbert space $\mathcal{H}$. As described in Section 4.2, a state ρ on $\mathcal{H}$ gives rise to an empirical model e on this compatibility structure. If the restriction of e to the sub-compatibility structure generated by a finite set of observables is contextual, then e is automatically contextual.

Next, we consider the situation where we have a natural transformation $t : T \stackrel{\cdot}{\longrightarrow} T'$ between the functor parts of commutative monads (T, η, μ, κ) and $(T', \eta', \mu', \kappa')$. Such a natural transformation converts empirical models for the monad T into empirical models for the monad T' by composition:

$$e : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E}) \quad \mapsto \quad t_{\mathcal{E}} \circ e : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T'(\mathcal{E}).$$

Proposition 9.3 (No-Go Descent Theorem) *Assume that $\mathbf{P}$ is simplicial. Let $t : T \stackrel{\cdot}{\longrightarrow} T'$ be a natural transformation, and $e : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$ an empirical model. If $t \circ e$ is contextual, so is e.*

Proof Suppose for a contradiction that e has a realization by a factorizable hidden-variable model. By the E = F Theorem 8.4, this implies that e extends to an empirical model e' defined on the full powerset of measurements $\mathbf{Q} = \mathcal{P}(\mathbf{A})$. Hence $t \circ e'$ is an extension of $t \circ e$ to $\mathbf{Q}$, which, by the E = F Theorem again, contradicts our assumption that $t \circ e$ is contextual. $\square$

9.3 Relational Collapse

We now apply the general notions of the previous sub-section to the case of distribution monads.

We say that a semiring is 0-**faithful** if it satisfies the following conditions:

$$\begin{array}{rcl} & 0 \neq 1 & \\ x + y = 0 & \Rightarrow & x = 0 \ \wedge \ y = 0 \\ x \cdot y = 0 & \Rightarrow & x = 0 \ \vee \ y = 0 \end{array}$$

Examples The semirings R of positive reals, and B of booleans, are both 0-faithful. Generalizing the case for R, the positive cone of any ordered integral domain is a 0-faithful semiring. Generalizing the case for B, any linear order with endpoints, viewed as an idempotent semiring, is 0-faithful.

A semiring homomorphism $h : R \to S$ is 0-**reflecting** if $h(x) = 0$ implies $x = 0$. Clearly, there is at most one 0-reflecting homomorphism $h_R : R \to \mathsf{B}$ to the Boolean semiring, defined by $h_R(0) = 0$, $h_R(x) = 1$ for $x \neq 0$.

Proposition 9.4 *A semiring R is 0-faithful if and only if $h_R : R \to \mathsf{B}$ is a homomorphism. Hence B is the terminal object in the category of 0-faithful semirings and 0-reflecting homomorphisms.*

We recall some general notions [14, 27]. A **morphism of commutative monads** $t : (T, \eta, \mu, \kappa) \to (T', \eta', \mu', \kappa')$ is a natural transformation $t : T \stackrel{.}{\longrightarrow} T'$ which commutes with the units, multiplications and commutative actions of the monads in the evident sense:

$$\eta' = t \circ \eta, \qquad \mu' \circ (t * t) = t \circ \mu, \qquad \kappa' \circ (t \times t) = t \circ \kappa.$$

Proposition 9.5 *A homomorphism of commutative semirings $h : R \to S$ induces a morphism of commutative monads $\bar{h} : \mathcal{D}_R \stackrel{.}{\longrightarrow} \mathcal{D}_S$:*

$$\bar{h}_X : \mathcal{D}_R(X) \to \mathcal{D}_S(X) :: d \mapsto h \circ d.$$

Note that, in the case of a 0-faithful semiring R, the natural transformation $\bar{h}$ maps an R-distribution to the characteristic function of its support.

We also note the following non-example.

Proposition 9.6 *There is no semiring homomorphism from the reals $(\mathbb{R}, +, 0, \cdot, 1)$ to the Booleans B or the positive reals R.*

Proof If there were a homomorphism h from the reals to B, we would have:

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

Since there is a homomorphism from the positive reals to the Booleans, this also shows that there cannot be one from the reals to the positive reals. $\square$

We say that a commutative monad T is **supported** if there is a natural transformation $t : T \stackrel{.}{\longrightarrow} \mathcal{D}_{\mathsf{B}}$ to the Boolean distribution monad, which we think of as the 'support' of the monad. Note that, by Propositions 9.4 and 9.5, any distribution monad over a 0-faithful semiring is supported.

If $e : \mathbf{1} \stackrel{.}{\longrightarrow} T(\mathcal{E})$ is an empirical model, the **relational collapse** of e is $t \circ e$.

The following is a special case of the No-Go Descent Theorem 9.3.

Proposition 9.7 (Relational Collapse Theorem) *Let T be a supported monad, and $e : \mathbf{1} \stackrel{.}{\longrightarrow} T(\mathcal{E})$ an empirical model. If the relational collapse of e is contextual, so is e.*

10 A Model-Specific No-Go Theorem

We shall now illustrate our approach by proving a Bell-type theorem; in fact, this will be a 'high-level' version of the Hardy construction [19]. The result is very general, and easily implies the usual formulations; we spell out how it applies to the quantum case.

We consider the following situation. We have two sites ('Alice' and 'Bob'). At each site two measurements can be performed: AX, AY and BX, BY respectively. Each measurement has two possible outcomes: U and D. The measurements at each site are incompatible, while as usual the measurements at different sites are compatible; thus this is a compatibility structure of the $\mathcal{C}_{\mathsf{L}}$ type, with maximal elements

$$\{\mathsf{AX}, \mathsf{BX}\}, \quad \{\mathsf{AX}, \mathsf{BY}\}, \quad \{\mathsf{AY}, \mathsf{BX}\}, \quad \{\mathsf{AY}, \mathsf{BY}\}.$$

Let T be a supported monad, with support $t : T \stackrel{.}{\longrightarrow} \mathcal{D}_{\mathsf{B}}$. We shall consider a class of empirical models $e : \mathbf{1} \stackrel{.}{\longrightarrow} T(\mathcal{E})$, which we shall call **Hardy models**. These will satisfy a constraint on their relational collapse $t \circ e$, expressed by the following table:

	(U, U)	(U, D)	(D, U)	(D, D)
{AX, BX}	1			
{AX, BY}	0			
{AY, BX}	0			
{AY, BY}				0

What this means is that the distribution $t \circ e$ (*i.e.* the support of e) must satisfy the indicated constraints on each maximal compatible set S; taking the value 1 for the outcome (U, U) on the first row, and 0 for the specified outcomes on the other three rows. The remaining values are left unspecified.

Recall that for compatibility structures of $\mathcal{C}_{\mathsf{L}}$ type, contextuality equates to the standard notion of non-locality.

Proposition 10.1 *Every Hardy model is contextual.*

Proof We shall show that for no Hardy model e can the relational collapse $t \circ e$ be extended to all sets of measurements. By the E = F Theorem 8.4, this will show that the relational collapse of e is contextual, and by the Relational Collapse Theorem 9.7, it follows that e is contextual.

For $S \in \mathcal{C}_{\mathsf{L}}$ and $s \in \mathcal{E}(S)$, we define $U(s) := \{s' \in O^M \mid s' \restriction S = s\}$. We shall name the sections with stipulated values for Hardy models:

$$
\begin{aligned}
s_0 &:= (\mathsf{AX}, \mathsf{BX}) \mapsto (\mathsf{U}, \mathsf{U}) \\
s_1 &:= (\mathsf{AX}, \mathsf{BY}) \mapsto (\mathsf{U}, \mathsf{U}) \\
s_2 &:= (\mathsf{AY}, \mathsf{BX}) \mapsto (\mathsf{U}, \mathsf{U}) \\
s_3 &:= (\mathsf{AY}, \mathsf{BY}) \mapsto (\mathsf{D}, \mathsf{D}).
\end{aligned}
$$

Suppose for a contradiction that $t \circ e$ can be extended to all subsets, in particular to M. By naturality (No-Signalling), we have:

$$
t \circ e_S(s) = \bigvee_{s' \in U(s)} t \circ e_M(s'). \tag{7}
$$

Thus if $t \circ e_S(s) = 0$, it must be the case, by the 0-faithfulness of the boolean semiring, that $t \circ e_M(s') = 0$ for all $s' \in U(s)$. From the three stipulated 0 entries for the Hardy model, we can conclude that for all $s \in U(s_1) \cup U(s_2) \cup U(s_3)$, we must have $t \circ e_M(s) = 0$.

If we write the sections in O^M in a compact form, as ordered lists of assignments of outcomes to the measurements $(\mathsf{AX}, \mathsf{BX}, \mathsf{AY}, \mathsf{BY})$, then $U(s_0)$ has four elements: UUUU, UUDU, UUUD, and UUDD. Of these, UUUU and UUDU are in $U(s_1)$, UUUD is in $U(s_2)$, and UUDD is in $U(s_3)$. Thus

$$
U(s_0) \;\subseteq\; U(s_1) \cup U(s_2) \cup U(s_3),
$$

and $t \circ e_M(s) = 0$ for all $s \in U(s_0)$. Using (7) again, this implies that $t \circ e_{\{\mathsf{AX},\mathsf{BX}\}}(s_0) = 0$, contradicting the stipulation that $t \circ e_{\{\mathsf{AX},\mathsf{BX}\}}(s_0) = 1$. □

Note that this proof makes no mention of hidden variables. It is entirely couched in terms of the empirical model, and extendability. For comparison, we shall also give a second, more standard proof, in terms of hidden variables.

Proof We assume for a contradiction that $t \circ e$ can be realized by a factorizable hidden-variable model h. By the Determinization Theorem 9.1, we can assume that h is in fact deterministic. We begin by using naturality (i.e., parameter independence) of h to infer some additional information about the model. From the fact that $t \circ e_{\{\mathsf{AX},\mathsf{BX}\}}(\mathsf{U}, \mathsf{U}) = 1$, we must have $h^{\lambda}_{\{\mathsf{AX},\mathsf{BX}\}}(\mathsf{U}, \mathsf{U}) = 1$ for some λ. By the naturality of h, and the 0-faithfulness of the boolean semiring, we must have

$h^{\lambda}_{\{\mathsf{AX}\}}(\mathsf{U}) = 1$. Using naturality of h again, we must have that for the compatible family $\{\mathsf{AX}, \mathsf{BY}\}$ (*i.e.* keeping Alice's measurement fixed but changing Bob's), some joint outcome whose Alice component is U must receive a non-zero value. Since we have stipulated that $t \circ e_{\{\mathsf{AX},\mathsf{BY}\}}(\mathsf{U}, \mathsf{U}) = 0$, we must have $h^{\lambda}_{\{\mathsf{AX},\mathsf{BY}\}}(\mathsf{U}, \mathsf{U}) = 0$, so the only remaining possibility is that $h^{\lambda}_{\{\mathsf{AX},\mathsf{BY}\}}(\mathsf{U}, \mathsf{D}) = 1$.

Exactly similar reasoning can then be used to infer successively that $h^{\lambda}_{\{\mathsf{AY},\mathsf{BY}\}}(\mathsf{U}, \mathsf{D}) = 1$, and $h^{\lambda}_{\{\mathsf{AY},\mathsf{BX}\}}(\mathsf{U}, \mathsf{D}) = 1$. The table now looks like this:

	(U,U)	(U,D)	(D,U)	(D,D)
$\{\mathsf{AX},\mathsf{BX}\}$	1			
$\{\mathsf{AX},\mathsf{BY}\}$	0	1		
$\{\mathsf{AY},\mathsf{BX}\}$	0	1		
$\{\mathsf{AY},\mathsf{BY}\}$		1		0

where it now indicates entries for h^{λ} rather than $t \circ e$. By naturality, from the first row we see that $h^{\lambda}_{\{\mathsf{BX}\}}(\mathsf{U}) = 1$, and from the third that $h^{\lambda}_{\{\mathsf{BX}\}}(\mathsf{D}) = 1$. But h is deterministic, yielding the required contradiction. $\square$

To show the application of this result, i.e., that quantum mechanics does give rise to contextual (in this case: non-local) behaviour, we must exhibit a quantum realization of a Hardy model.

A detailed discussion of physical realizations is given in [19, 33]. We shall just give a simple concrete instance.

We consider the two-qubit system, with AY and BY measurement in the computational basis. We take $\mathsf{U} = 0$, $\mathsf{D} = 1$. The eigenvectors for AX are taken to be

$$\sqrt{\frac{3}{5}}|0\rangle + \sqrt{\frac{2}{5}}|1\rangle, \qquad -\sqrt{\frac{2}{5}}|0\rangle + \sqrt{\frac{3}{5}}|1\rangle$$

and similarly for BX. The state is taken to be

$$\sqrt{\frac{3}{8}}|10\rangle \; + \; \sqrt{\frac{3}{8}}|01\rangle \; - \; \frac{1}{2}|00\rangle.$$

One can then calculate the probabilities to be

$$p_{\{\mathsf{AX},\mathsf{BY}\}}(\mathsf{U},\mathsf{U}) \;=\; p_{\{\mathsf{AY},\mathsf{BX}\}}(\mathsf{U},\mathsf{U}) \;=\; p_{\{\mathsf{AY},\mathsf{BY}\}}(\mathsf{D},\mathsf{D}) = 0,$$

and $p_{\{\mathsf{AX},\mathsf{BX}\}}(\mathsf{U},\mathsf{U}) = 0.09$, which is very near the maximum attainable value [33].

11 Strong Contextuality

It is clear that factorizability is a strong notion. By weakening it, we can consider stronger forms of No-Go theorems. The Kochen-Specker theorem [26] is a result of exactly this form. Our aim in the next two sections is to elucidate these ideas in our general setting.

We shall henceforth work under the assumption that $\mathbf{P}$ is simplicial. We take it to be of the $\mathbf{P}_{\mathcal{C}}$, where $\mathcal{C}$ is a finite compatibility structure.

Suppose that $e : \mathbf{1} \stackrel{\cdot}{\longrightarrow} T(\mathcal{E})$ is an empirical model for a monad T, and that T is **supported** in the sense defined in Section 9.3, i.e., there is a natural transformation $t : T \stackrel{\cdot}{\longrightarrow} \mathcal{D}_{\mathsf{B}}$ to the boolean distribution monad. In particular, any distribution monad $\mathcal{D}_R$ over a 0-faithful semiring R has such a natural transformation, by Propositions 9.4 and 9.5.

For each $S \in \mathcal{C}$, we define $\mathcal{S}_e(S) := \mathsf{supp}(t_{\mathcal{E}(S)}(e_S)) \subseteq \mathcal{E}(S)$.

Proposition 11.1 *$\mathcal{S}_e$ is a sub-presheaf of $\mathcal{E}$.*

Proof Suppose that $s \in \mathcal{S}_e(S)$, and $S' \subseteq S$. We must show that $s|S' \in \mathcal{S}(S')$. Firstly, $s \in \mathcal{S}_e(S)$ means that $t_{\mathcal{E}(S)}(e_S)(s) = 1$. Now from naturality of e and t, we have:

$$\begin{tikzcd}
\mathbf{1} \arrow[r, "e_S"] \arrow[dr, "e_{S'}"'] & T\mathcal{E}(S) \arrow[r, "t_{\mathcal{E}(S)}"] \arrow[d, "T\mathcal{E}(\iota_{S',S})"] & \mathcal{D}_{\mathsf{B}}\mathcal{E}(S) \arrow[d, "\mathcal{D}_{\mathsf{B}}\mathcal{E}(\iota_{S',S})"] \\
 & T\mathcal{E}(S') \arrow[r, "t_{\mathcal{E}(S')}"'] & \mathcal{D}_{\mathsf{B}}\mathcal{E}(S')
\end{tikzcd}$$

This yields

$$t_{\mathcal{E}(S')}(e_{S'})(s|S') \;=\; t_{\mathcal{E}(S)}(e_S)|S'(s|S') \;=\; \bigvee_{s'|S'=s|S'} t_{\mathcal{E}(S)}(e_S)(s').$$

Thus $t_{\mathcal{E}(S)}(e_S)(s) = 1$ implies that $t_{\mathcal{E}(S')}(e_{S'})(s|S') = 1$, and $s|S' \in \mathcal{S}(S')$ as required. □

The following observation is key. We say that an empirical model e is **non-contextual** if it has a realization by a factorizable hidden-variable model.

Proposition 11.2 *If e is non-contextual, then $\mathcal{S}_e$ has a global section.*

Proof By Proposition 8.3, e extends to a model $e' : \mathbf{1} \dot{\longrightarrow} T(\mathcal{E}')$, where $\mathcal{E}'$ is the canonical extension of $\mathcal{E}$ to $\mathcal{P}(M)$. We can similarly extend $\mathcal{S}_e$ to $\mathcal{S}_{e'} \hookrightarrow \mathcal{E}'$ by $\mathcal{S}_{e'}(S) := \mathsf{supp}(t_{\mathcal{E}(S)}(e'_S))$. Now $\mathcal{S}_{e'}(M) = \mathsf{supp}(t_{\mathcal{E}(M)}(e'_M))$ will contain some section s. We can define $x_S := s|S$ for each $S \subseteq M$. The argument given in the previous Proposition shows that $x_S \in \mathcal{S}_{e'}(S)$. Hence x is a global section of $\mathcal{S}_{e'}$, and its restriction to $\mathcal{C}$ is a global section of $\mathcal{S}_e$. □

We say that an empirical model e is **strongly contextual** if $\mathcal{S}_e$ has no global section. By the previous Proposition, strong contextuality implies contextuality.

It will be useful to have a simple equivalent condition for the existence of a global section. We consider the case of a simplicial poset $\mathbf{P}$, which we can take to be a compatibility structure $\mathcal{C}$ over a finite set M. We consider sub-presheaves $\mathcal{S}$ of $\mathcal{E}_O$. We shall write $\mathsf{Max}(\mathcal{C})$ for the maximal elements of $\mathcal{C}$.

Proposition 11.3 *The following are equivalent:*

- *$\mathcal{S}$ has a global section.*
- *There is a function $f : M \to O$ such that, for all $S \in \mathsf{Max}(\mathcal{C})$, $f|S \in \mathcal{S}(S)$.*

Proof If $x : \mathbf{1} \dot{\longrightarrow} \mathcal{S}$ is a global section, the values $x_{\{m\}}$, $m \in M$, define a function $f : M \to O$. For any $S \in \mathsf{Max}(\mathcal{C})$ and $m \in S$, naturality of x implies that $x_S|\{m\} = x_{\{m\}}$, so $f|S = x_S \in \mathcal{S}(S)$.

For the converse, given f, we define $x_S = f|S \in \mathcal{S}(S)$. Given $S \in \mathcal{C}$, for some $S' \in \mathsf{Max}(\mathcal{C})$, $S \subseteq S'$. Since $f|S' \in \mathcal{S}(S')$, $x_S = f|S = (f|S')|S \in \mathcal{S}(S)$, by naturality of the inclusion $i : \mathcal{S} \hookrightarrow \mathcal{E}$. □

Proposition 11.4 *Strong contextuality is strictly stronger than contextuality. There are empirical models arising from quantum systems which are contextual, but not strongly contextual.*

Proof We consider the quantum system giving rise to a Hardy model which we described in Section 10. If we compute the support over all outcomes of maximal compatible measurements explicitly, we obtain the following table.

	(U, U)	(U, D)	(D, U)	(D, D)
{AX, BX}	1	1	1	1
{AX, BY}	0	1	1	1
{AY, BX}	0	1	1	1
{AY, BY}	1	1	1	0

Now consider the following assignment:

$$\mathsf{AX} \mapsto \mathsf{D}, \quad \mathsf{AY} \mapsto \mathsf{U}, \quad \mathsf{BX} \mapsto \mathsf{D}, \quad \mathsf{BY} \mapsto \mathsf{U}.$$

We can check that in every context S, the section arising by restricting this mapping to the measurements in S is in the support. For example, the section at $\{\mathsf{AX}, \mathsf{BX}\}$ is (D, D). Thus this assignment defines a global section. $\square$

We shall now observe a surprising difference between the Hardy model, and the well-known GHZ model [18, 17]. Both are, of course, famously used to give proofs of non-locality, which in our framework appears as a special case of contextuality. However, as we have just seen, the Hardy model does have a global section in its support, and so does not exhibit strong contextuality. As we shall now see, by contrast the GHZ model **is** strongly contextual. Moreover, a simplified version of the standard argument to prove non-locality [32] can be used to show that it satisfies this stronger property.

The GHZ model is a tripartite system with two measurements at each part, each with two possible outcomes. There are no compatibility relations within each part, so this yields a compatibility structure of $\mathcal{C}_{\mathsf{L}}$ type. We shall use the following notation: 122 will indicate the (maximal) context in which measurement 1 is selected at the first site, and measurement 2 at the second and third sites. We shall refer to the possible outcomes for each measurement as R or G, and write, in the context where some measurements ijk have been selected, RRG for the section which assigns outcome R to i and j, and outcome G to k.

The class of GHZ models are defined to be those whose support satisfies the following conditions. Let $P = \{122, 212, 221\}$. Then we require that for all $p \in P$, the support of the model at p is $\{RRR, RGG, GRG, GGR\}$, and that the support at 111 is $\{RRG, RGR, GRR, GGG\}$. There are well-known quantum realizations of such models, using spin measurements, and the GHZ state

$$\frac{|000\rangle + |111\rangle}{\sqrt{2}}.$$

Proposition 11.5 *GHZ models are strongly contextual.*

Proof Let us assume for a contradiction that the GHZ model e has a global section, induced by a map $f : M \to O$. We know that we must have $f|S \in \{RRR, RGG, GRG, GGR\}$, for $S \in P$. Since 122 and 221 have the same middle measurement, if $f|122 \in \{RRR, GRG\}$, then $f|221 \in \{RRR, GRG\}$, and if $f|122 \in \{RGG, GGR\}$, then $f|221 \in \{RGG, GGR\}$. Thus there are 8 possible joint assignments by the global section to 122 and 221.

We can now check that any of these completely determines the assignment to 212. Suppose for example that $f|122 = RRR$, $f|221 = GRG$. Then $f|212 = G{-}R$, and the only consistent possibility for the middle outcome is that $212 \mapsto GGR$. We can represent this joint assignment to the measurements in P as the 'instruction'

$$RGG$$
$$GRR$$

The rubric is that the i'th row gives the assignment by the global section when the measurements are set to i, $i = 1, 2$. A similar analysis applies to the other seven cases.

We can tabulate this well-known 'Mermin instruction set' [32] as follows:

$$\begin{array}{cccc} RRR & RGG & GRG & GGR \\ RRR & RGG & GRG & GGR \\ \\ RGG & RRR & GGR & GRG \\ GRR & GGG & RRG & RGR \end{array}$$

For any of these cases, consider the measurement setting 111. Since each top row contains an odd number of R's, and the possible outcomes for 111 all contain an even number of R's, we obtain the desired contradiction. $\square$

12 A Generalized Kochen-Specker Theorem

The notion of strong contextuality depends only on the support of an empirical model; as the results of the previous section suggest, the obstructions to the existence of a global section follow from the compatibility structure, and more precisely from the existence of incompatible measurements.

This leads us to the following idea. Consider a sub-presheaf $i : \mathcal{S} \overset{\cdot}{\hookrightarrow} \mathcal{E}$, and models $e : \mathbf{1} \overset{\cdot}{\longrightarrow} T(\mathcal{E})$ for a supported monad T which are **based on** $\mathcal{S}$, *i.e.* whose support factors through this inclusion:

$$\mathcal{S}_e \overset{\cdot}{\hookrightarrow} \mathcal{S} \overset{\cdot}{\hookrightarrow} \mathcal{E}.$$

Now we have the the following.

Proposition 12.1 *If $\mathcal{S}$ has no global sections, then any empirical model based on $\mathcal{S}$ must be strongly contextual.*

Proof If the support $\mathcal{S}_e$ of a model e based on $\mathcal{S}$ had a global section, so would $\mathcal{S}$. $\square$

Note that this formulation is **model-independent**; the property is question is determined purely by the $\mathbf{P}$-(pre)sheaf $\mathcal{S}$. This corresponds to the state-independent nature of the Kochen-Specker theorem.

We shall refer to such a situation as exhibiting **generic strong contextuality**.

Proposition 12.2 *If $\mathcal{S}$ is a sheaf with $\mathcal{S}(S) \neq \varnothing$ for all $S \in \mathsf{Max}(\mathcal{C})$, it is determined by its restriction to $\mathsf{Max}(\mathcal{C})$. Given any $S \in \mathcal{C}$:*

$$s \in \mathcal{S}(S) \iff \exists S' \in \mathsf{Max}(\mathcal{C}), s' \in \mathcal{S}(S').\, s'|S = s. \tag{8}$$

Proof If $\mathcal{S}$ is a presheaf, $S' \in \mathsf{Max}(\mathcal{C})$, and $s' \in \mathcal{S}(S')$, then we must have $s'|S \in \mathcal{S}(S)$ whenever $S \subseteq S'$. Conversely, if $s \in \mathcal{S}(S)$, for some $S' \in \mathsf{Max}(\mathcal{C})$, $S \subseteq S'$. Since $\mathcal{S}(S') \neq \varnothing$, there is some $s' \in \mathcal{S}(S')$, and $s'' = s'|S'' \in \mathcal{S}(S'')$, where $S'' = S' \setminus S$. Now $\{S, S''\}$ form a disjoint cover of S', and s, s'' form a compatible family of sections, so since $\mathcal{S}$ is a sheaf, there is $s_0 \in \mathcal{S}(S')$ with $s_0|S = s$. $\square$

12.1 Equivalent Conditions for Generic Strong Contextuality

We now turn to the issue of proving generic strong contextuality for various classes of presheaves.

The Finite Case

We consider the case where $\mathcal{E} = \mathcal{E}_{O_k}$, where $O_k = \{1, \ldots, k\}$. In this case, there is a pleasing confluence of sheaf-theoretic, combinatorial, and computational questions.

We recall that a **constraint satisfaction problem** (CSP) [35] 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 (S, C), where $S \subseteq V$, and $C \subseteq K^S$. (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 $s : V \to K$ **satisfies** a constraint (S, C) if $s \restriction S \in C$. A **solution** of the CSP $(V, K, \mathcal{R})$ is an assignment $s : V \to K$ which satisfies every constraint in $\mathcal{R}$.

Now suppose we are given a presheaf $\mathcal{S} \xhookrightarrow{\cdot} \mathcal{E}_{O_k}$. We can define the corresponding CSP as $(M, O_k, \{\mathcal{S}(S) \mid S \in \mathsf{Max}(\mathcal{C})\})$.

Proposition 12.3 *The presheaf $\mathcal{S}$ has a global section if and only if the corresponding CSP has a solution.*

Proof This is just a rewording of Proposition 11.3. $\square$

The Boolean Case

We now specialize further, to the case where $O = O_2$.

In this case, we interpret the two values as truth values, and M as a set of propositional variables. Given $X \subseteq \mathcal{E}_{O_2}(S)$, we write $\mathsf{DNF}(X)$ for the formula

$$\bigvee_{s \in X} \Big(\bigwedge_{m \in S, s(m) = \mathsf{true}} m \ \wedge \bigwedge_{m \in S, s(m) = \mathsf{false}} \neg m \Big).$$

Proposition 12.4 *The presheaf $\mathcal{S}$ has a global section if and only if the formula*

$$\bigwedge_{S \in \mathsf{Max}(\mathcal{C})} \mathsf{DNF}(\mathcal{S}(S))$$

has a satisfying assignment.

Proof Again, this follows directly from Proposition 11.3. $\square$

12.2 The Quantum Case

We now wish to apply these ideas to the quantum case. Until now, we have been taking the presheaf $\mathcal{S}$ as given, and seeing how the existence of a global section for $\mathcal{S}$ can be formulated equivalently as a natural algorithmic problem. Now we wish to take advantage of the special features of the quantum situation to **define** a specific form of presheaf, such that we know *a priori* that **all** quantum states give rise to empirical models which live on this presheaf. We can then apply results on non-existence of global sections for these presheaves directly to show the generic strong contextuality of quantum mechanics — and of any other theory which gives rise to similar structures.

Note the striking contrast between a result of this form — essentially the classical Kochen-Specker theorem [26] — and model-dependent results. The latter exploit features of specific quantum states, notably entanglement, in order to show contextuality or non-locality. By contrast, the generic type of result shows that, for various choices of finite sets of measurements, all quantum states, whether entangled or not, will display strongly contextual behaviour.

We shall focus on **dichotomic quantum observables**, as discussed in Sections 3.4 and 4.2. We shall consider a particular form of dichotomic observables, which will be convenient for our purposes. Given unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_k$ in a Hilbert space $\mathcal{H}$, we write

$$A_{\mathbf{e}_i} := i \cdot \mathbf{P}_{e_i} + 0 \cdot \mathbf{P}^{\perp}_{e_i}.$$

Then we can take $M = \{A_{\mathbf{e}_1}, \ldots, A_{\mathbf{e}_k}\}$ 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

$$\sum_{i \in I} i \cdot \mathbf{P}_{e_i} \; + \; 0 \cdot \mathbf{P}_{\{\mathbf{e}_i | i \in I\}^{\perp}}. \tag{9}$$

If we measure any state with this observable, the outcome must be that we get exactly one of the branches $\mathbf{P}_{e_i}$, 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 $\mathcal{E}_O$, 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 m_i as a label for $A_{\mathbf{e}_i}$, and taking $s(m_i) = 1$ for the outcome corresponding to $\mathbf{P}_{e_i}$ 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 in the compatibility structure — exactly one measurement must be assigned 1.

This leads to the following general definition of a presheaf for any finite compatibility structure $\mathcal{C}$. Given a set $S \in \mathcal{C}$, we write $\mathsf{ONE}(S)$ for the formula

$$\bigvee_{m \in S} (m \; \wedge \bigwedge_{m' \in S \setminus \{m\}} \neg m').$$

We define an assignment $\mathcal{Q}$ on the maximal elements of $\mathcal{C}$:

$$\mathcal{Q}(S) := \{s : S \to \{0,1\} \mid s \models \mathsf{ONE}(S)\}, \qquad S \in \mathsf{Max}(\mathcal{C}).$$

Clearly $\mathcal{Q}(S) \neq \varnothing$ for all $S \in \mathsf{Max}(\mathcal{C})$.

We then extend this assignment to all $S \in \mathcal{C}$, as in (8):

$$s \in \mathcal{Q}(S) \iff \exists S' \in \mathsf{Max}(\mathcal{C}), s' \in \mathcal{S}(S').\, s' | S = s.$$

It is easy to see that, for non-maximal $S \in \mathcal{C}$, s is in $\mathcal{Q}(S)$ if and only if it sets at most one measurement to 1.

Proposition 12.5 *$\mathcal{Q}$ is a presheaf. It is not a sheaf.*

Proof It is straightforward to verify the naturality condition. To see that $\mathcal{Q}$ is not a sheaf, consider a partition $\{S_1, S_2\}$ of $S \in \mathsf{Max}(\mathcal{C})$. We can find elements $s_i \in \mathcal{Q}(S_i)$, which satisfy $\mathsf{ONE}(S_i)$, $i = 1, 2$; since they set distinct elements of S to 1, they have no common extension to $\mathcal{Q}(S)$. □

As another straightforward consequence of Proposition 11.3, we have:

Proposition 12.6 *The presheaf $\mathcal{Q}$ has a global section if and only if there is a satisfying assignment for the formula*

$$\bigwedge_{S \in \mathsf{Max}(\mathcal{C})} \mathsf{ONE}(S).$$

We shall now give a simple combinatorial condition on the compatibility structure $\mathcal{C}$ which is implied by the existence of a global section for $\mathcal{Q}$ defined on $\mathcal{C}$. Violation of this condition therefore suffices to prove that no global section exists.

For each $m \in M$, we define

$$\mathcal{C}(m) := \{S \in \mathsf{Max}(\mathcal{C}) \mid m \in S\}.$$

Proposition 12.7 *If $\mathcal{Q}$ has a global section, then every common divisor of $\{|\mathcal{C}(m)| \mid m \in M\}$ must divide $|\mathsf{Max}(\mathcal{C})|$.*

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

$$|\mathsf{Max}(\mathcal{C})| = \sum_{m \in M'} |\mathcal{C}(m)|.$$

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

For example, if every $m \in M$ appears in an even number of maximal elements of $\mathcal{C}$, while $\mathcal{C}$ has an odd number of maximal elements, then $\mathcal{S}$ has no global section.

The simplest example of this situation is the 'triangle', *i.e.* the compatibility structure with maximal elements

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

An example where M has 18 elements, and there are 9 maximal compatible sets, each with four elements, such that each element of M is in two of these, appears in the 18-vector proof of the Kochen-Specker Theorem in [10].

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 compatible family is an orthogonal set of vectors. We shall now put this kind of construction in a broader graph-theoretical setting.

12.3 Connections to Graph Theory

The compatibility structures which arise from quantum systems are of a particular form: they are generated by a **symmetric binary compatibility relation**, namely, that the observables pairwise commute. Thus, for example, the 'triangle' cannot arise from quantum observables.

This suggests that we should take account of 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 cliques of G forms a compatibility structure $\mathcal{C}_G$.

Let G be a graph. A set $S \subseteq V_G$ is called a **stable transversal** [5] if for every maximal clique C of G (*i.e.* for every $C \in \mathsf{Max}(\mathcal{C}_G)$), $|S \cap C| = 1$. 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 12.8 *Let G be a graph. The presheaf $\mathcal{Q}$ defined on $\mathcal{C}_G$ has a global section if and only if G has a stable transversal.*

Proof If $\mathcal{Q}$ has a global section, then by Proposition 11.3, there is a function $f : M \to \{0,1\}$ with $f|C \in \mathcal{Q}(C)$ for every $C \in \mathsf{Max}(\mathcal{C}_G)$. This means that $f|C \models \mathsf{ONE}(C)$. Thus if we define $T := \{m \in M \mid f(m) = 1\}$, then T is a stable traversal of G.

Conversely, suppose that T is a stable transversal of G. If we define f as the characteristic function of T on M, then $f|C \models \mathsf{ONE}(C)$ for each maximal clique C of G, and so $f|C \in \mathcal{Q}(C)$, and by Proposition 11.3, $\mathcal{Q}$ has a global section. □

There are results in the graph theory literature giving sufficient conditions for the existence of stable transversals. For example, we have the following from [25].

Theorem 12.9 *If a graph G satisfies $\omega(G) > 2/3(\Delta(G)+1)$, then G has a stable transversal.*

Here $\omega(G)$ is the maximum clique size, and $\Delta(G)$ the maximum degree of any vertex. Note that in the construction in [10], $\omega(G) = 4$, and $\Delta(G) = 6$. It is shown in [25] that this bound is tight.

In order to apply graph-theoretic results to the quantum situation, we need to know which graphs can arise from families of quantum observables; specifically, from the form of dichotomic observables we are considering, satisfying (9). Recall that $A_{\mathbf{e}_i}$ commutes with $A_{\mathbf{e}_j}$ if and only if $\mathbf{e}_i$ is orthogonal to $\mathbf{e}_j$. Thus, we are interested in graphs which 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 [30], so we shall say that such graphs have a **faithful orthogonal co-representation** in $\mathbb{R}^d$. In order to preserve the intended interpretation of $\mathcal{C}_G$, and hence the validity of the constraints expressed by $\mathcal{Q}$, we must 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 compatibility structure $\mathcal{C}_G$ such that the presheaf $\mathcal{Q}$ defined on $\mathcal{C}_G$ has no global section; and every such graph can be realized by dichotomic observables in the way we have described. 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 [10], as well as the various sets of 31 or more vectors which have been found in $\mathbb{R}^3$ [26, 9, 38, 37].

A final desideratum is to provide a purely graph-theoretic condition for the existence of a faithful orthogonal co-representation. In [30, 29] the following result is proved.

Theorem 12.10 *Every graph on n nodes whose complementary graph is $(n-d)$-connected has a faithful orthogonal co-representation in $\mathbb{R}^d$.*

12.4 The Kochen-Specker Theorem and the Spectral Presheaf

We shall now relate our notion of generic strong contextuality to the original Kochen-Specker theorem [26], as expressed in the language of presheaves in [21].

Let $\mathcal{H}$ be a Hilbert space. Following [21, 13], we define the poset $\mathbf{C}(\mathcal{H})$ of all **commuting subalgebras** of $\mathbf{B}(\mathcal{H})$, the algebra of bounded linear operators on $\mathcal{H}$, ordered by inclusion.

The central object of study in the 'topos approach' [13] to quantum foundations is the **spectral presheaf** $\Sigma(\mathcal{H})$. This is defined on the poset $\mathbf{C}(\mathcal{H})$, and assigns to each commutative algebra A the **Gelfand spectrum** of A, whose points are the **characters** of A, *i.e.* the algebra homomorphisms from A to the complex numbers. The functorial action is by restriction.

The spectral presheaf provides an elegant reformulation of the Kochen-Specker notion of **partial algebra** [26]. The point of their paper was to show that there are finitary obstructions to the existence of any homomorphism from the partial algebra of observables to a commutative algebra; and hence to the existence of (deterministic) hidden variables for quantum mechanics. In terms of the spectral presheaf, as observed in [21], this translates into the non-existence of a global section for this presheaf.

We now show that the generic strong contextuality notion we have been studying easily implies this result. We fix some finite-dimensional subspace $\mathcal{H}_0$ of $\mathcal{H}$. Let M be the set of all dichotomic observables on $\mathcal{H}_0$ of the form $A_e = a \cdot \mathbf{P}_e + 0 \cdot \mathbf{P}_e^{\perp}$, as studied in Section 12.2. We can form the compatibility structure $\mathcal{C}$ of all finite sets of commuting observables in M, and hence the poset $\mathbf{P}_{\mathcal{C}}$.

There is a natural monotone map $\phi : \mathbf{P}_{\mathcal{C}} \to \mathbf{C}(\mathcal{H}_0)$ which sends a set S of commuting observables to the commutative subalgebra $\bar{S}$ it generates in $\mathbf{B}(\mathcal{H}_0)$. We can compose $\Sigma(\mathcal{H}_0)$, defined on $\mathbf{C}(\mathcal{H}_0)$, with ϕ to get a presheaf on $\mathbf{P}_{\mathcal{C}}$, which we denote by $\Sigma_{\mathbf{P}}$.

Lemma 12.11 *There is a presheaf morphism $u : \Sigma_{\mathbf{P}} \overset{\cdot}{\longrightarrow} \mathcal{Q}$.*

Proof Given $S \in \mathcal{C}$, and a character $c : \bar{S} \to \mathbb{C}$, we must define $u_S(c) \in \mathcal{E}(S)$, and verify that it satisfies $\mathsf{ONE}(S)$. Since c is an unital algebra homomorphism, we have, for each $A_e = a \cdot \mathbf{P}_e + 0 \cdot \mathbf{P}_e^{\perp}$ in S:

$$c(\mathbf{P}_e) = c(\mathbf{P}_e^2) = c(\mathbf{P}_e)^2 \in \{0, 1\}, \qquad c(\mathbf{P}_e + \mathbf{P}_e^{\perp}) = c(I) = 1.$$

Thus, for each $A_e \in S$, c assigns 1 or 0 to $\mathbf{P}_e$, and the complementary value to $\mathbf{P}_e^{\perp}$. We define a section $s = u_S(c) : S \to \{0, 1\}$, by $s(A_e) := c(\mathbf{P}_e)$. For any resolution of the identity $\sum_i \mathbf{P}_i = I$, c assigns 1 to exactly one of the P_i, and 0 to the others; hence s is in $\mathcal{Q}(S)$. Since Σ acts by restriction, this assignment is natural in S, and thus defines a presheaf morphism u. □

Proposition 12.12 *If $\mathcal{Q}$ has no global sections, then $\Sigma(\mathcal{H})$ has no global sections.*

Proof If $\Sigma(\mathcal{H}_0)$ had a global section $x : \mathbf{1} \overset{\cdot}{\longrightarrow} \Sigma(\mathcal{H})$, we could define a global section $y : \mathbf{1} \overset{\cdot}{\longrightarrow} \Sigma_{\mathbf{P}}$ by $y_S := x_{\bar{S}} : \mathbf{1} \to \Sigma(\mathcal{H})(\bar{S}) = \Sigma_{\mathbf{P}}(S)$. Then by Lemma 12.11, there would be a global section $u \circ y : \mathbf{1} \overset{\cdot}{\longrightarrow} \mathcal{Q}$, contra hypothesis. Since a global section of $\Sigma(\mathcal{H})$ would restrict to a global section of $\Sigma(\mathcal{H}_0)$, this result extends to spectral presheaves on infinite-dimensional spaces. □

Moreover, we note that by the Restriction Theorem 9.2, it suffices to study finite compatibility structures which can be embedded into $\mathcal{C}$, such as those generated by Kochen-Specker graphs, as described in Section 12.3.

13 Concluding Remarks

The development in this paper has been carried out at a high level of generality and abstraction, relative to the usual discussions of foundations of quantum mechanics. Nevertheless, from the mathematical point of view, several directions for possible further generalizations suggest themselves.

- Firstly, many of our results have used the hypothesis that the poset $\mathbf{P}$ is simplicial. It would be of interest to see how far this could be weakened.
- The finiteness assumption incorporated in working with simplicial posets goes along with the generally discrete setting of this paper. Working in an enriched setting, with some topological or measure-theoretic structure in the background, might allow more general forms of probabilistic model and of measurement structure to be considered. This could enable results such as those in [7] to be generalized to our setting.
- We have used commutative monads throughout. It may be interesting to use **strong monads** [27], which would allow for non-commutative effects [34].

References

[1] S. Abramsky. Relational Hidden Variables and Non-Locality. *Arxiv preprint arXiv:1007.2754*, 2010. Submitted for publication.

[2] S. Abramsky and N. Tzevelekos. Introduction to categories and categorical logic. In B. Coecke, editor, *New Structures for Physics*, volume 813 of *Lecture Notes in Physics*, pages 3–94. Springer, 2010.

[3] S. Awodey. *Category theory.* Oxford University Press, 2010.

[4] J.S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195–200, 1964.

[5] C. Berge and P. Duchet. Strongly perfect graphs. *Topics on perfect graphs*, 21:57–61, 1984.

[6] A. Björner. Posets, regular CW complexes and Bruhat order. *European J. Combin*, 5(1):7–16, 1984.

[7] A. Brandenburger and H.J. Keisler. Observable implications of unobservable variables. Available at `http://pages.stern.nyu.edu/∼abranden/oiuv-06-22-10.pdf`, April 2010.

[8] A. Brandenburger and N. Yanofsky. A classification of hidden-variable properties. *Journal of Physics A: Mathematical and Theoretical*, 41:425302, 2008.

[9] J. Bub. Schütte's tautology and the Kochen-Specker theorem. *Foundations of Physics*, 26(6):787–806, 1996.

[10] A. Cabello, J.M. Estebaranz, and G. García-Alcaine. Bell-Kochen-Specker theorem: A proof with 18 vectors. *Physics Letters A*, 212(4):183–187, 1996.

[11] G. Cattani and G. Winskel. Presheaf models for concurrency. In *Computer Science Logic*, pages 58–75. Springer, 1997.

[12] W.M. Dickson. *Quantum chance and non-locality.* Cambridge University Press, 1999.

[13] A. Döring and C.J. Isham. A topos foundation for theories of physics: II. Daseinisation and the liberation of quantum theory. *Journal of Mathematical Physics*, 49:053516, 2008.

[14] S. Eilenberg and J.C. Moore. Adjoint functors and triples. *Illinois Journal of Mathematics*, 9(3):381–398, 1965.

[15] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48(5):291–295, 1982.

[16] G.C. Ghirardi, A. Rimini, and T. Weber. A general argument against superluminal transmission through the quantum mechanical measurement process. *Lettere Al Nuovo Cimento (1971–1985)*, 27(10):293–298, 1980.

[17] D.M. Greenberger, M.A. Horne, A. Shimony, and A. Zeilinger. Bell's theorem without inequalities. *American Journal of Physics*, 58:1131, 1990.

[18] D.M. Greenberger, M.A. Horne, and A. Zeilinger. Going beyond Bell's theorem. In M. Kafatos, editor, *Bell's theorem, Quantum theory, and Conceptions of the Universe*, pages 69–72. Kluwer, 1989.

[19] L. Hardy. Nonlocality for two particles without inequalities for almost all entangled states. *Physical Review Letters*, 71(11):1665–1668, 1993.

[20] C. Heunen, N.P. Landsman, and B. Spitters. A topos for algebraic quantum theory. *Communications in mathematical physics*, 291(1):63–110, 2009.

[21] C.J. Isham and J. Butterfield. Topos perspective on the Kochen-Specker theorem: I. Quantum states as generalized valuations. *International Journal of Theoretical Physics*, 37(11):2669–2733, 1998.

[22] B. Jacobs. Semantics of weakening and contraction. *Annals of Pure and Applied Logic*, 69(1):73–106, 1994.

[23] J.P. Jarrett. On the physical significance of the locality conditions in the Bell arguments. *Noûs*, 18(4):569–589, 1984.

[24] J.B. Kennedy. On the empirical foundations of the quantum no-signalling proofs. *Philosophy of Science*, 62(4):543–560, 1995.

[25] A.D. King. Hitting all maximum cliques with a stable set. *Arxiv preprint arXiv:0911.1741*, 2009.

[26] S. Kochen and E.P. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17(1):59–87, 1967.

[27] A. Kock. Monads on symmetric monoidal closed categories. *Archiv der Mathematik*, 21(1):1–10, 1970.

[28] A. Kock. Bilinearity and cartesian closed monads. *Math. Scand*, 29:161–174, 1971.

[29] L. Lovász, M. Saks, and A. Schrijver. A correction: orthogonal representations and connectivity of graphs. *Linear Algebra and its Applications*, 313:101–106, 2000.

[30] M. Lovász, M. Saks, and A. Schrijver. Orthogonal representations and connectivity of graphs. *Linear Algebra and its Applications*, 114:439–454, 1989.

[31] S. Mac Lane and I. Moerdijk. *Sheaves in geometry and logic: A first introduction to topos theory*. Springer, 1992.

[32] N.D. Mermin. Quantum mysteries revisited. *Am. J. Phys*, 58(8):731–734, 1990.

[33] N.D. Mermin. Quantum mysteries refined. *American Journal of Physics*, 62:880, 1994.

[34] E. Moggi. Notions of computation and monads. *Information and computation*, 93(1):55–92, 1991.

[35] U. Montanari. Networks of constraints: Fundamental properties and applications to picture processing 1. *Information sciences*, 7:95–132, 1974.

[36] F.J. Oles. Type categories, functor categories and block structure. In M. Nivat amd J.C. Reynolds, editor, *Algebraic Semantics*, pages 543–574. Cambridge University Press, 1985.

[37] M. Pavičić, J.P. Merlet, B. McKay, and N.D. Megill. Kochen–Specker vectors. *Journal of Physics A: Mathematical and General*, 38:1577, 2005.

[38] A. Peres. Two simple proofs of the Kochen-Specker theorem. *Journal of Physics A: Mathematical and General*, 24:L175, 1991.

[39] B.C. Pierce. *Basic category theory for computer scientists*. The MIT Press, 1991.

[40] A. Shimony. Events and processes in the quantum world. *Quantum concepts in space and time*, pages 182–203, 1986.

[41] R. Tennent. Functor-category semantics of programming languages and logics. *Category Theory and Computer Programming*, pages 206–224, 1986.