Various Equivalent Formalisms for Contextuality

Ming Yin

Abstract

Contextuality is a property of empirical models that shows considerable promise in our attempt to better understand the properties of quantum mechanics that allow certain quantum algorithms to have a computational advantage over classical computation. In this review, we focus on three recent formalisms of contextuality, via sheaf theory, logical inequalities and graph theory. We study examples of quantum systems that exhibit different strengths of contextuality and give equivalent proofs using these different frameworks. We then provide an overview of some interesting results that have been proved about the role of contextuality in quantum advantage, and suggest promising further research directions.

1 Introduction

Despite numerous examples of quantum algorithms believed to offer a (in many cases exponential) speed-up over classical computation, we lack a general understanding of the precise features of quantum mechanics that admit quantum computational advantage. Considerable research has been undertaken on various properties, such as superposition [1], entanglement [2, 3, 4], discord [5] and contextuality. In this essay, we focus on this last concept and detail three ways of defining and studying contextuality: via sheaf theory, logical inequalities and graph theory.

Section 2 puts the study of contextuality in context by giving a brief overview of consequential work over the last few decades. Section 3 introduces *empirical models* and provides three important examples. Sections 4 and 5 describe the formalisms in detail and prove key results about contextual empirical models. We explain how these formalisms, which at first seem very different, in fact lead to equivalent descriptions of contextuality and thus can be unified. Section 6 then explores different strengths of contextuality and the numerous equivalent ways that each strength can be described or proved. Finally, in Sections 7 and 8 we give an overview of current and potential future applications of contextuality to better understanding quantum advantage.

2 History

Contextuality has its origins in a 1935 argument of Einstein, Podolsky, and Rosen (EPR) [6], that quantum theory is an 'incomplete' description of reality, in the sense that any complete description should, at least theoretically, allow us to determine with certainty the value of any measurable property. In other words, there should exist underlying *local hidden variables*, even if in practice they cannot be observed. Einstein had been particularly motivated by his concern about the nature of quantum mechanics that allows 'spooky action at a distance' and was skeptical that the theory truly admitted an intrinsic description of the universe. The EPR paper inspired a great deal of discussion about whether quantum scenarios could be explained by local hidden variable theories which, according to a certain probability distribution, admit a global assignment of measurement outcomes to all possible measurements.

A landmark paper by Bell in 1964 [7] gave the first counterexample, showing by means of the violation of an inequality that there are some quantum-realizable scenarios that cannot be explained by *any* local hidden variable theory. This was the birth of the concept of *non-locality*, which countless future works elaborated on, e.g. [8]. The term *contextuality* was later coined to generalize this notion to scenarios that permit non-local measurements, including single-qubit scenarios. The 1967 Kochen–Specker theorem [9] demonstrated the inherent contextual nature of quantum mechanics by showing that no non-contextual hidden variable theory can reproduce the predictions of quantum mechanics whenever the dimension of the Hilbert space is at least three. In the 2011 paper by Abramsky and Brandenburger [10], the various ideas and proofs about contextuality that had been floating around up until that point were formalized using sheaf theory, in such a way as to be fundamentally independent of quantum theory.

3 Empirical models

Here, our basic setup will follow the approach of Abramsky and Brandenburger [10], which we will examine in more detail in Section 4.1. Given an experimental scenario, let $\mathcal{M}$ denote the set of all measurements that can be performed, and let $\mathcal{O}$ denote the set of possible outcomes of each measurement. Let $\mathcal{C} \subset \mathcal{P}(\mathcal{M})$ be a collection of *contexts* $C \subseteq \mathcal{M}$; each context is a set of measurements that can be jointly performed. Usually we require that the contexts are *maximal* and *cover* all measurements, i.e. $\bigcup \mathcal{C} = \mathcal{M}$. For each subset of measurements $U \subseteq \mathcal{M}$, the collection of all possible (formal) joint outcomes can be expressed as the set of all functions mapping each measurement in U to an outcome, i.e. $\mathcal{O}^U = \{s : U \to \mathcal{O}\}$. Each $s \in \mathcal{O}^U$ can be thought of as a *formal event*.

In particular, when the subset of $\mathcal{M}$ is a context C, the formal events are physically permissible. Hence we will simply call these *events*. Given a collection of formal events $\mathcal{O}^U$, a *probability distribution* on $\mathcal{O}^U$ is a function $d : \mathcal{O}^U \to \mathbb{R}_{\geq 0}$ such that $\sum_{s\in\mathcal{O}^U} d(s) = 1$. To construct an *empirical model*, we associate to each context C a particular distribution d_C. Note that if $U' \subseteq U \subseteq \mathcal{M}$ and $s \in \mathcal{O}^U$ then there is a natural restriction $s|_{U'} : U' \to \mathcal{O}$. Given a distribution $d_U : \mathcal{O}^U \to \mathbb{R}_{\geq 0}$, we define the restriction of d_U to U' as the function $d_U|_{U'} : U' \to \mathcal{O}$ with action

$$s' \mapsto \sum_{\substack{s\in\mathcal{O}^U,\\ s|_{U'}=s'}} d_U(s). \tag{3.1}$$

Intuitively, this is consistent with how we marginalize a joint probability distribution over all possible outcomes of one part.

In the context of quantum systems, given a Hilbert space $\mathcal{H}$, elements of $\mathcal{M}$ label observables on $\mathcal{H}$. The contexts label maximal commuting subsets of observables. A quantum state ρ on $\mathcal{H}$ represents an empirical model where each probability distribution d_C is given by $d_C(s) = \mathrm{Tr}(\rho\Pi_s)$, where Π_s is a rank-1 projector onto the state that results from the event s.

Now we introduce three types of empirical models that are instructive in demonstrating different strengths of contextuality. We will return to these examples in later sections.

3.1 Bell models

A *Bell model* with parameters (n, k, l) is a scenario with n spatially separated sites, k possible measurements at each site and l possible outcomes for each measurement. An important example is the following $(2, 2, 2)$ model [11] that can be derived from Bell's original paper [7]. Denote the two sites by A and B. Denote the possible measurements at A by a and a', and similarly for B. The possible outcomes are 0 and 1. Thus we have $\mathcal{M} = \{a, a', b, b'\}$ and $\mathcal{O} = \{0, 1\}$. A valid set of measurements has precisely one measurement at each site, so $\mathcal{C} = \{a, a'\} \times \{b, b'\}$. If we construct a table where rows are labelled by contexts and columns are labelled by possible joint outcomes, each cell then corresponds to an event. The probability distribution on each $\mathcal{O}^C$, $C \in \mathcal{C}$, is given by the values of the cells in the corresponding row. See Table 3.1 for the completed probability table. Each row has been labelled with the set of events that are represented by the cells in that row. For example, the top-left cell corresponds to the event $s \in \mathcal{O}^{\{a,b\}}$ where $s(a) = 0$, $s(b) = 0$.

We make an important observation: regardless of the measurement context, the *marginal probability* that the measurement at A has outcome 0 is always $1/2$; similarly, the marginal probability that the measurement at B has outcome 0 is always $1/2$. This illustrates a necessary property of all empirical models: for all $C, C' \in \mathcal{C}$, the distributions d_C and $d_{C'}$ must be *compatible*, i.e.

$$d_C|_{C\cap C'} = d_{C'}|_{C\cap C'}, \tag{3.2}$$

A	B	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$	
a	b	$1/2$	0	0	$1/2$	$\mathcal{O}^{\{a,b\}}$
a	b'	$3/8$	$1/8$	$1/8$	$3/8$	$\mathcal{O}^{\{a,b'\}}$
a'	b	$3/8$	$1/8$	$1/8$	$3/8$	$\mathcal{O}^{\{a',b\}}$
a'	b'	$1/8$	$3/8$	$3/8$	$1/8$	$\mathcal{O}^{\{a',b'\}}$

Table 3.1: A $(2,2,2)$ Bell model

where the restriction of a distribution is defined as in Equation 3.1. For Bell models, this compatibility property is equivalent to no-signalling: roughly, the measurement at one site cannot affect the probability distribution over outcomes at other sites.

The $(2,2,2)$ model above is an example of a quantum Bell model. In general, quantum Bell models can be represented by an underlying composite space $\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_n$ and local observables $A_i^j = I \otimes \cdots \otimes \underbrace{A_i}_{j\text{th site}} \otimes \cdots \otimes I$. This particular model can be realized by a Bell state $(|00\rangle + |11\rangle)/\sqrt{2}$ with local spin measurements on each qubit that are in the XY-plane of the Bloch sphere and at a relative angle of $\pi/3$ [12].

We will see in Section 4.1.1 that this $(2,2,2)$ Bell model exhibits contextuality. In fact, since we have a partitioned system with local measurements, it is a special case of contextuality known as *non-locality*. This was historically the first example of contextuality.

3.2 Hardy model

Hardy devised a quantum model involving two qubits and three well-chosen bases [13]. As with the $(2,2,2)$ Bell model above, a local measurement is performed on each qubit. The precise probability distributions depend on the choice of quantum state, but here, we consider the *possibilistic* version of the Hardy model, where we take a standard Hardy model and replace all positive probabilities with 1. This results in Table 3.2. It can be thought of as the *support* of the standard model.

A	B	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
a	b	1	1	1	1
a	b'	0	1	1	1
a'	b	0	1	1	1
a'	b'	1	1	1	0

Table 3.2: The possibilistic Hardy model

In this case, rather than probability distributions that map $\mathcal{O}^U$ into $\mathbb{R}_{\geq 0}$, we have 'distributions' $d : \mathcal{O}^U \to \mathbb{B} = \{0,1\}$, the Booleans. If we identify 'addition' with the OR operator $\vee$, then the distributions as given by Table 3.2 still satisfy the compatibility property (3.2).

Later, we will see that the possibilistic Hardy model exhibits a somewhat stronger form of contextuality than the Bell model above, but not as strong as GHZ models.

3.3 GHZ models

Greenberger, Horne and Zeilinger [14] introduced these models and used them to demonstrate an even stronger form of contextuality [15]. For n qubits, the permissible measurements are local Pauli X and Y spin measurements. For X, the eigenstates $(|0\rangle + |1\rangle)/\sqrt{2}$ and $(|0\rangle - |1\rangle)/\sqrt{2}$ correspond to measurement outcomes 0 and 1 respectively. For Y, the eigenstates $(|0\rangle + \mathrm{i}\,|1\rangle)/\sqrt{2}$ and $(|0\rangle - \mathrm{i}\,|1\rangle)/\sqrt{2}$ correspond to measurement outcomes 0 and 1 respectively. We prepare the system in the GHZ state, written in the computational basis as

$$\frac{|0\cdots 0\rangle + |1\cdots 1\rangle}{\sqrt{2}}.$$

	000	001	010	011	100	101	110	111
XXX	1/4	0	0	1/4	0	1/4	1/4	0
XXY	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
XYX	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
XYY	0	1/4	1/4	0	1/4	0	0	1/4
YXX	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
YXY	0	1/4	1/4	0	1/4	0	0	1/4
YYX	0	1/4	1/4	0	1/4	0	0	1/4
YYY	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

Table 3.3: The three-qubit GHZ model

We then perform a local X or Y measurement on each qubit.

It is easy to check that the probability table for a GHZ model has a support which is characterized by the following properties [10]. Let C_i be the context associated to the ith row of the table, and let y_i be the number of Y measurements in C_i.

- If y_i is odd, the ith row has full support.
- If $y_i \equiv 0 \pmod 4$, the ith row is supported on joint outcomes that have an even number of 1s.
- If $y_i \equiv 2 \pmod 4$, the ith row is supported on joint outcomes that have an odd number of 1s.

We also have that, for each row, the distribution is uniform on the support. For example, Table 3.3 shows the probabilities for the GHZ model on three qubits.

4 Approaches to contextuality

Now we will see how each of the above models exhibits contextuality. We begin by fleshing out the sheaf-theoretic approach of Abramsky and Brandenburger [10]. Then we introduce an alternative, graph-theoretic approach due to Cabello, Severini and Winter [16]. Over the next few sections of this review, we will describe equivalent ways of expressing contextual properties using both formalisms, which provide us with an abundance of tools for studying contextual models.

4.1 The sheaf-theoretic approach

The following formalism requires a very basic understanding of *categories*. Very roughly, a category is a collection of objects and arrows between the objects. Important examples include **Set**, where the objects are sets and arrows are functions between sets; and partially ordered sets (posets) $(P, \leq)$, where the objects are elements of P and there is a single arrow from p to q if $p \leq q$, with no other arrows. Every category has identity arrows id_A that map an object A to itself. If $\mathcal{A}$ and $\mathcal{B}$ are categories, a *functor* $F : \mathcal{A} \to \mathcal{B}$ maps every object in $\mathcal{A}$ to an object in $\mathcal{B}$, and also maps every arrow in $\mathcal{A}$ to an arrow in $\mathcal{B}$ in a way that preserves composition and identities.

Recall that we defined a formal event to be a function $s \in \mathcal{O}^U$, where $U \subseteq \mathcal{M}$. In the language of sheaf theory, functions $U \to \mathcal{O}$ are known as *sections* over U, so that $\mathcal{O}^U$ is the set of sections over U. Note that the power set $\mathcal{P}(\mathcal{M})$ is a poset with partial order $\subseteq$. We can turn $\mathcal{P}(\mathcal{M})$ into a *category* by drawing a single arrow from U' to U if $U' \subseteq U$. We denote by $\mathcal{P}(\mathcal{M})^{\mathsf{op}}$ the *opposite category*, in which the arrows are reversed. Consider the assignment $\mathcal{E}$ that maps a subset of measurements U to its set of sections $\mathcal{O}^U$. Furthermore, if $U' \subseteq U$, we have the natural restriction map $\mathcal{O}^U \to \mathcal{O}^{U'}$, $s \mapsto s|_{U'}$. This action induces a natural map from arrows in $\mathcal{P}(\mathcal{M})^{\mathsf{op}}$ to arrows in $\{\mathcal{O}^U \mid U \subseteq \mathcal{M}\} \subset$ **Set**. (Note that we need the opposite category as the restriction action is 'arrow reversing'.) Thus we have a functor $\mathcal{E} : \mathcal{P}(\mathcal{M})^{\mathsf{op}} \to$ **Set**, which is known as a *presheaf*.

In order to promote a presheaf to a *sheaf*, the following *gluing condition* must hold, which is indeed the case for $\mathcal{E}$. Let $U \subseteq \mathcal{M}$ and suppose we have a cover of U, denoted by $\{U_i\}_{i\in I}$. Suppose further that we have a family of sections $\{s_i \in \mathcal{E}(U_i)\}_{i\in I}$, one for each U_i, which is compatible, i.e. $s_i|_{U_i\cap U_j} = s_j|_{U_i\cap U_j}$ for all $i, j \in I$. Then there is a unique section $s \in \mathcal{E}(U)$ such that $s|_{U_i} = s_i$ for all $i \in I$. In fact, $\mathcal{E}$ trivially satisfies this *sheaf condition* since the domain of any section is discrete.

In Section 3.2 we saw an example of a model that uses an alternative to probability distributions. We now formalize this idea. A *commutative semiring* $(R, +, 0, \cdot, 1)$ is a generalization of a commutative ring where we do not require the existence of additive identities. For example, the non-negative reals form a semiring $(\mathbb{R}_{\geq 0}, +, 0, \cdot, 1)$. Other important examples are (trivially) the reals $\mathbb{R}$ and the Booleans, $\mathbb{B} = (\{0,1\}, \vee, 0, \wedge, 1)$, where $\vee$ is the logical OR and $\wedge$ is the logical AND.

Let R be a semiring. Given a set X, an R-*distribution* on X is a function $d : X \to R$ such that d has finite support and $\sum_{x\in X} d(x) = 1$. Thus R-distributions are a generalization of probability distributions, which occur in the special case where $R = \mathbb{R}_{\geq 0}$. Denote by $\mathcal{D}_R(X)$ the set of R-distributions on X. We can elevate $\mathcal{D}_R$ to a functor by considering the following action on functions. Given a function $f : X \to Y$, let $\mathcal{D}_R(f) : \mathcal{D}_R(X) \to \mathcal{D}_R(Y)$ be the function with mapping

$$d \mapsto \left(y \mapsto \sum_{\substack{x\in X,\\ f(x)=y}} d(x) \right). \tag{4.1}$$

Then we have a functor $\mathcal{D}_R$: **Set** $\to$ **Set**.

Now we can form the presheaf $\mathcal{D}_R\mathcal{E} : \mathcal{P}(\mathcal{M})^{\mathsf{op}} \to$ **Set**, which maps each measurement subset U to the set of distributions on $\mathcal{O}^U$. Also, given $U' \subseteq U \subseteq \mathcal{M}$, so that there is an arrow from U' to U in $\mathcal{P}(\mathcal{M})$, $\mathcal{D}_R\mathcal{E}$ maps this arrow to the arrow $\mathcal{D}_R\mathcal{E}(U) \to \mathcal{D}_R\mathcal{E}(U')$, which is a function between sets. Let $d \in \mathcal{D}_R\mathcal{E}(U)$. From (4.1), we see that d is mapped to the distribution in $\mathcal{D}_R\mathcal{E}(U') = \mathcal{D}_R(\mathcal{O}^{U'})$ that is given by

$$s' \mapsto \sum_{\substack{s\in\mathcal{O}^U,\\ s|_{U'}=s'}} d(s).$$

This is exactly the marginalization property (3.1)! Thus it makes sense to denote this distribution by $d|_{U'}$, the restriction of d to U'.

Lastly, we formalize the definition of an empirical model. Given a set of measurements $\mathcal{M}$, suppose we have a collection $\mathcal{C}$ of contexts that cover $\mathcal{M}$. Conventionally, we impose the condition that $\mathcal{C}$ contains no nested sets, i.e. if $C', C \in \mathcal{C}$ and $C' \subseteq C$ then $C' = C$. Then an *empirical model* Ω is a family of distributions, exactly one for each context,

$$\Omega = \{d_C \in \mathcal{D}_R\mathcal{E}(C) \mid C \in \mathcal{C}\},$$

such that the distributions are compatible. (3.2) is an equivalent statement of this compatibility.

4.1.1 Contextuality as the absence of a global section

We saw above that $\mathcal{E}$ is always a sheaf, but for $\mathcal{D}_R\mathcal{E}$ the situation is more complicated. Recall that the sheaf condition requires that a gluing condition hold for any $U \subseteq \mathcal{M}$ and cover $\{U_i\}$. Instead, we restrict ourselves to asking whether a *given empirical model* $\Omega = \{d_C\}$ satisfies the sheaf condition with respect to the cover $\mathcal{C}$. Since we know that the d_C are already compatible with each other, this amounts to asking whether there exists a *global section* $d \in \mathcal{D}_R\mathcal{E}(\mathcal{M}) = \mathcal{D}_R(\mathcal{O}^{\mathcal{M}})$ that is compatible with this family of distributions, i.e. $d|_C = d_C$ for all $C \in \mathcal{C}$. Recall that d is a distribution over all global sections in $\mathcal{O}^{\mathcal{M}}$. We will call elements of $\mathcal{O}^{\mathcal{M}}$ *global assignments* from now on in order to avoid confusion.

It is important to note that these global assignments correspond to *canonical hidden variables*, since every measurement in $\mathcal{M}$ is assigned an outcome. If we take $R = \mathbb{R}_{\geq 0}$ and return to the familiar land of probability distributions, we are thus precisely asking whether there is a hidden variable model, i.e. a probability distribution over hidden variables $\mathcal{M} \to \mathcal{O}$, that is compatible with Ω.

Therefore, we finally have a formal sheaf-theoretic definition of contextuality.

Definition 4.1. *An empirical model Ω defined on the presheaf $\mathcal{D}_R\mathcal{E}$, with measurements $\mathcal{M}$, contexts $\mathcal{C}$ and outputs $\mathcal{O}$, is* non-contextual *if it has a compatible global section $d \in \mathcal{D}_R\mathcal{E}(\mathcal{M})$. Otherwise, Ω is* contextual.

We return to the Bell model from Section 3.1. Abramsky and Brandenburger (AB) use the following argument to show that the model has no global section. We fix an enumeration (C, s) of the contexts C and the sections s over C, as well as an enumeration (t) of all the global assignments $t \in \mathcal{O}^{\mathcal{M}}$. Then we can define an *incidence matrix* $\mathbf{M}$ with rows indexed by (C, s), columns indexed by t, and entries

$$\mathbf{M}[(C,s),t] = \begin{cases} 1 & \text{if } t|_C = s, \\ 0 & \text{otherwise.} \end{cases}$$

Each column corresponds to a deterministic model $\delta^t \in \mathcal{D}_R\mathcal{E}(\mathcal{M})$ with global assignment t (i.e. $\delta^t(t) = 1$, $\delta^t(t') = 0$ for all $t' \neq t$), where the positions of the 1s tell us the result of restricting t to each context [12]. Thus each global section d compatible with Ω can be represented by a vector in the column span of $\mathbf{M}$. We can express this notationally as $d = \sum_{t \in \mathcal{O}^{\mathcal{M}}} p_t \delta^t$ for some $p_t \in R$, which means that for all $C \in \mathcal{C}$

$$d|_C = \sum_{t \in \mathcal{O}^{\mathcal{M}}} p_t \delta^t|_C.$$

In order to ensure that d is indeed a distribution, we require that $\sum_t p_t = 1$. Now, an empirical model $\Omega = \{d_C\}$ can be identified with a vector $\mathbf{v}$ of length $|(C, s)|$ with entries $\mathbf{v}[(C, s)] = d_C(s)$. Putting everything together, we have the following result.

Proposition 4.1. *Ω has a compatible global section if and only if the system $\mathbf{Mx} = \mathbf{v}$ has a solution over R such that the sum of the entries of $\mathbf{x}$ is 1.*

For our $(2, 2, 2)$ Bell model, there are 4 contexts and 4 possible joint measurements. Thus $\mathbf{M}$ has 16 rows. Since $|\mathcal{M}| = 4$, we can label the global assignments by length-4 bit strings $0000, 0001, \ldots$ corresponding to the possible global joint outcomes. The full matrix for this model is shown in Figure 4.1. The sections are ordered left to right, top to bottom, according to Table 3.1. Each length-4 bit string corresponds to a global assignment with respect to the order $aba'b'$. With this linear-algebraic formulation of the problem, it can easily be proved that a solution $\mathbf{x}$ over $\mathbb{R}_{\geq 0}$ does not exist (Proposition 4.2 of [10]).

In fact, since the possibilistic Hardy model (Section 3.2) has the exact same measurements and contexts as this Bell model, the same $\mathbf{M}$ can be used to show that the Hardy model is also contextual. This time, we view the system $\mathbf{Mx} = \mathbf{v}$ as being over the Booleans $\mathbb{B}$ so that, for example, the first entry of $\mathbf{Mx}$ is the expression $x_1 \vee x_2 \vee x_3 \vee x_4$, where $\mathbf{x} = (x_1, x_2, \ldots)$.

We note here that AB further show, using a more sophisticated linear-algebraic argument, that if we loosen the semiring R to the reals $\mathbb{R}$, then in fact *any* empirical model Ω over any measurement cover has a global section (Theorem 5.5 of [10]). This is consistent with the long-hypothesized idea that all quantum mechanical systems *can* be described by hidden variable models if we allow 'negative probabilities'[(i)] [18].

4.2 CSW inequalities

Cabello, Severini and Winter (CSW) used graph theory to form inequalities that determine the existence of contextuality. A *simple, undirected graph* $G = (V, E)$ consists of a set of *vertices* $V = \{v_1, \ldots, v_n\}$ and *edges* $E \subset V \times V$ such that $(v, v) \notin E$ for any $v \in V$ and $(v, v') \in E$ if and only if $(v', v) \in E$. We say that two

[(i)]Note however that there are contextual models that cannot be realized by a quantum system, such as the PR box [17, 10].

$$
\begin{array}{r}
\\
\{a \mapsto 0,\, b \mapsto 0\} \\ \\ \\ \\
\{a \mapsto 0,\, b' \mapsto 0\} \\ \\ \\ \\
\{a' \mapsto 0,\, b \mapsto 0\} \\ \\ \\ \\
\{a' \mapsto 0,\, b' \mapsto 0\} \\ \\ \\ \\
\end{array}
\begin{array}{c}
\begin{array}{cccccccccccccccc}
0000 & 0001 & 0010 & 0011 & \cdots & & & & & & & & & & &
\end{array} \\
\begin{bmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1
\end{bmatrix}
\end{array}
$$

Figure 4.1: The matrix $\mathbf{M}$ for our $(2, 2, 2)$ Bell model and the possibilistic Hardy model

vertices v, v' are *adjacent* if there is an edge connecting them, i.e $(v, v') \in E$. A *weighted graph* (G, w) has an associated function $w : V \to \mathbb{R}_{\geq 0}$ that assigns a weight to each vertex [19].

Given an experimental scenario, we form an *exclusivity graph*, in which the vertices represent (formal) events and two different vertices are connected by an edge if and only if they represent mutually exclusive events. For example, a *classical model* on a classical sample space Λ can be represented by a graph G and probabilities p_i, where each vertex v_i corresponds to an event $e_i \subset \Lambda$, adjacent vertices correspond to *disjoint* events, and there is a probability distribution on Λ such that the probability of the event e_i is p_i. Similarly, a *quantum model* on a Hilbert space $\mathcal{H}$ in a pure state $|\psi\rangle$ can be represented by a graph G and probabilites p_i, where each vertex v_i corresponds to a projector Π_i and adjacent vertices correspond to *orthogonal* projectors. The p_i then give the probability of obtaining the measurement outcome associated to each projector: $p_i = \langle\psi| \Pi_i |\psi\rangle$.

It is helpful to introduce the following definitions. An *independent set* of vertices consists of vertices none of which are pairwise adjacent, i.e. the vertices correspond to consistent events that could occur simultaneously. A *clique* consists of vertices that are all pairwise adjacent. The *independence degree* of a vertex v is the number of vertices that are not adjacent to v. The *independence number* of a graph is the maximum independence degree over all vertices, or equivalently the largest size of an independent set. The *weighted independence number* of a weighted graph (G, w), denoted by $\alpha(G, w)$, is the maximum value of the sum $\sum_{v \in I} w(v)$ where I is an independent set. CSW showed that, given a classical model with graph G and probabilities p_i, any weight function w on the vertices of G satisfies

$$\sum_i w(v_i) p_i \leq \alpha(G, w), \tag{4.2}$$

and there exists a choice of w that achieves the upper bound. Therefore, if the p_i violate this inequality for some w, then they cannot arise from a classical model and hence the model is contextual.

Definition 4.2. *An empirical model Ω represented by a graph G and probabilities p_i is* contextual *if it violates a CSW inequality (4.2) for some weight function w.*

As we will see in Section 5.1, this definition is equivalent to Definition 4.1.

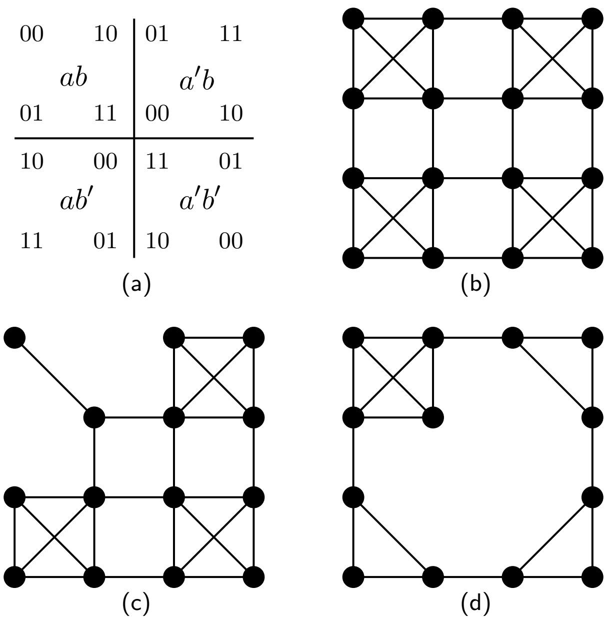

Figure 4.2: Using the labelling given by (a), we can construct a general exclusivity graph $G(\mathcal{M}, \mathcal{C})$ (b) for the measurements and contexts shared by the $(2, 2, 2)$ Bell model and the possibilistic Hardy model. The support graphs for the Bell model (c) and Hardy model (d) are also shown.

4.3 Bridging these approaches

De Silva showed an elegant way to construct an exclusivity graph from an empirical model as defined by AB[(ii)] [19]. Given an experimental scenario with measurements $\mathcal{M}$ and contexts $\mathcal{C}$, the vertices of the exclusivity graph $G(\mathcal{M}, \mathcal{C})$ correspond to sections (events) $s : C \to \mathcal{O}$, where $C \in \mathcal{C}$. Two vertices $s_1 : C_1 \to \mathcal{O}$ and $s_2 : C_2 \to \mathcal{O}$ are adjacent if s_1 and s_2 are mutually exclusive, i.e. if $s_1|_{C_1 \cap C_2} \neq s_2|_{C_1 \cap C_2}$. Note that two different events over the same context are always mutually exclusive. Therefore, any independent set of G can have at most one vertex for each context, so the maximum size of an independent set is $|\mathcal{C}|$, the number of contexts.

Given an empirical model $\Omega = \{d_C\}$, the *support graph* of Ω is the subgraph $G_\Omega(\mathcal{M}, \mathcal{C})$ obtained by keeping only the events that are possible, i.e. events s such that $d_C(s) > 0$ (where $s \in \mathcal{O}^C$). The probability associated to each vertex $s \in \mathcal{O}^{\mathcal{C}}$ is $p(s) = d_C(s)$.

Examples of an exclusivity graph and support graphs for the $(2, 2, 2)$ Bell model and the possibilistic Hardy model can be found in Figure 4.2.

With this construction, we have the following equivalence involving global assignments/hidden variables (Lemma 1 of [19]).

Proposition 4.2. *Given an experimental scenario with measurements $\mathcal{M}$ and contexts $\mathcal{C}$, there is a bijection between canonical hidden variables $t \in \mathcal{O}^{\mathcal{M}}$ and independent sets $I \subset G(\mathcal{M}, \mathcal{C})$ of size $|\mathcal{C}|$.*

Proof. Using the reasoning above, an independent set I of size $|\mathcal{C}|$ has exactly one event over each context. Since the set is independent, the events are consistent, i.e. they agree on common measurements. Thus we can map the set to a global assignment $t : \mathcal{M} \to \mathcal{O}$ as the contexts cover $\mathcal{M}$.

Conversely, given a global assignment t, the vertices given by restricting t to each context, $\{t|_C : C \to \mathcal{O} \mid C \in \mathcal{C}\}$, are mutually consistent are thus form an independent set of size $|\mathcal{C}|$. It is straightforward to see that these two maps are inverses of one another. □

[(ii)] We usually take the semiring R to be $\mathbb{R}_{\geq 0}$.

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

Table 5.1: The $(2,2,2)$ Bell model with particular events shaded

5 Logical Bell inequalities

So far, we have seen how we can use either the sheaf-theoretic formalism or CSW inequalities to show that a model is contextual. In fact, inequalities have been used to show non-locality, and later contextuality, since Bell's original paper [7]. Abramsky and Hardy developed the idea of *logical Bell inequalities*, which formalize the numerous contextuality-related inequalities that had previously been constructed.

We use notation from propositional logic and consider the case where $\mathcal{O} = \{0,1\}$[iii]. We can then form a propositional formula for each event over a context, with Boolean variables labelled by the measurements in that context [12, 19]. For example, referring to the Bell model in Table 3.1, the events corresponding to the first row can be expressed as the following formulas.

$$\neg a \wedge \neg b, \quad \neg a \wedge b, \quad a \wedge \neg b, \quad a \wedge b.$$

In general, given an event $s \in \mathcal{O}^C$, the corresponding formula is

$$\varphi_s = \bigwedge_{m \in C} \begin{cases} m & \text{if } s(m) = 1, \\ \neg m & \text{if } s(m) = 0. \end{cases}$$

Furthermore, given a context C, we can form a formula by choosing a subset of events $S \subset \mathcal{O}^C$ and defining

$$\varphi_S = \bigvee_{s \in S} \varphi_s,$$

which is satisfied if and only if at least one of the event formulas φ_s is satisfied.

Now, we derive the following general result using elementary probability theory. Suppose we have propositional formulas $\varphi_1, \ldots, \varphi_N$ with respective truth probabilities $p_1, \ldots, p_N$. Let $\Phi = \bigwedge_i \varphi_i$ and let P be the probability of Φ. Then we have

$$\begin{aligned} 1 - P = p(\neg\Phi) &= p\left(\bigvee_i \neg\varphi_i\right) \\ &\leq \sum_i p(\neg\varphi_i) = N - \sum_i p_i, \end{aligned}$$

where $p_i = p(\varphi_i)$. This gives us $\sum_i p_i - P \leq N - 1$. Now, if the formulas φ_i are not correlated, then $P = 0$, in which case we have $\sum_i p_i \leq N - 1$. This is an example of a logical Bell inequality.

Consider once again our $(2,2,2)$ Bell model. In particular, using the shaded events in Table 5.1, we can form the following formulas, one for each row (i.e. context) of the table:

$$\begin{aligned} \varphi_1 &= (\neg a \wedge \neg b) \vee (a \wedge b) &&= a \leftrightarrow b, \\ \varphi_2 &= (\neg a \wedge \neg b') \vee (a \wedge b') &&= a \leftrightarrow b', \\ \varphi_3 &= (\neg a' \wedge \neg b) \vee (a' \wedge b) &&= a' \leftrightarrow b, \\ \varphi_4 &= (\neg a' \wedge b') \vee (a' \wedge \neg b) &&= a' \oplus b'. \end{aligned}$$

[iii] It is easy to generalize this to other sets of outcomes; see Section VIII of [12].

Here, $\leftrightarrow$ means if and only if (so $a \leftrightarrow b$ means a and b have the same logical value), and $\oplus$ is the logical XOR (or equivalently addition modulo 2). The probability of φ_1 is $p_1 = 1/2 + 1/2 = 1$, and so on. It is easy to check that, if the first three formulas are correlated, then φ_4 is FALSE. Thus $\Phi = \bigwedge_{i=1}^{4} \varphi_i$ is impossible. Since $N = 4$, the associated Bell inequality is

$$\sum_{i=1}^{4} p_i \leq 3.$$

However, using the probabilities in Table 5.1, the sum of the left-hand side is $13/4$, so this model actually violates the Bell inequality by $1/4$. This is again evidence of contextuality: in the words of Abramsky and Hardy (AH), 'We simply cannot regard the variables as each representing a global, context-independent quantity.' This corresponds to the idea of the non-existence of a global section, as previously discussed.

In general, suppose that a model has contexts $\mathcal{C}$ and that $C_1, \ldots, C_N$ is an enumeration of some contexts; note that they are not necessarily distinct. Then the associated logical Bell inequality has the form

$$\sum_{i=1}^{N} k_i p(\varphi_i) \leq K, \tag{5.1}$$

where the k_i are non-negative integers and each φ_i is a formula whose variables are drawn from C_i. The k_i can be thought of as weights associated to the formulas φ_i. K is a positive integer such that any *multiset* of formulas φ_i, where φ_i appears at most k_i times, that has size larger than K is not completely satisfied by any global assignment $\mathcal{M} \to \mathcal{O}$. AH showed that these inequalities completely characterize contextuality given any experimental scenario (Theorem VI.5 of [12]). An empirical model is contextual if and only if it violates at least one associated logical Bell inequality.

5.1 Logical Bell inequalities and exclusivity graphs

We notice similarities between CSW inequalities (4.2) and logical Bell inequalities (5.1). De Silva formalized this observation for Bell inequalities of the form $\sum_i p_i \leq N - 1$ (Theorem 1 of [19]). Here, we flesh out the details that generalize this result to general logical Bell inequalities.

Theorem 5.1. *Logical Bell inequalities are derived from CSW inequalities on the exclusivity graph $G(\mathcal{M}, \mathcal{C})$.*

Proof. Suppose we have an empirical model with measurements $\mathcal{M}$ and contexts $\mathcal{C}$. Let $C_1, \ldots, C_N$ be an enumeration of some contexts and let $\varphi_1, \ldots, \varphi_N$ be propositional formulas on these contexts. Now suppose the logical Bell inequality derived from these formulas is

$$\sum_{i=1}^{N} k_i p(\varphi_i) \leq K.$$

For each φ_i, let $S_i \subset C_i$ be the set of events $s : C_i \to \mathcal{O}$ that appear in φ_i (the *support* of φ_i). Each event $s_{ij} \in S_i$ is represented by a vertex in the graph $G(\mathcal{M}, \mathcal{C})$. Note that, for each i, $\sum_j p(s_{ij}) = p(\varphi_i)$. Thus we can rewrite the Bell inequality as

$$\sum_i \sum_j k_i p(s_{ij}) \leq K.$$

Define a weight function w on the graph such that $w(s_{ij}) = k_i$ for each s_{ij} in each S_i and $w(s) = 0$ if s is not in any of the supports. Then the CSW inequality in this case is

$$\sum_{i,j} k_i p(s_{ij}) \leq \alpha(G, w).$$

We claim that $\alpha(G, w)$ cannot be greater than K. Indeed, suppose it is. Since any independent set can have at most one event over each context, it follows that there exists an independent set $\{s_{ij_i}\}_{i\in I}$ such that $\sum_{i\in I} w(s_{ij_i}) = \sum_{i\in I} k_i > K$. By Proposition 4.2, there is a global assignment that is compatible with these events, so the formulas φ_i, $i \in I$ are correlated. Then the multiset

$$\{\varphi_i^{k_i}\}_{i\in I},$$

where φ_i appears k_i times, is logically consistent. But this multiset has size $\sum_{i\in I} k_i > K$, which is a contradiction. □

So, to summarize, so far we have seen three equivalent characterizations of contextuality:

- the absence of a global section with respect to the distribution presheaf $\mathcal{D}_R\mathcal{E}$,
- the violation of a CSW inequality associated to a weighted exclusivity graph,
- and the violation of a logical Bell inequality.

6 Strengths of contextuality

6.1 Logical contextuality

We return to the possibilistic Hardy model from Section 3.2. In Section 4.1.1 we discussed a linear-algebraic way to prove contextuality by showing the non-existence of solutions to a matrix equation. We can use a similar method for the Hardy model, except we note that this time the semiring is simply $\mathbb{B}$ rather than $\mathbb{R}_{\geq 0}$. In general, given any empirical model over the semiring $\mathbb{R}_{\geq 0}$, if the corresponding system $\mathbf{Mx} = \mathbf{v}$ has a solution, then the resulting system after mapping each entry into $\mathbb{B}$ by the semiring homomorphism

$$\begin{aligned} \mathbb{R}_{\geq 0} &\to \mathbb{B} \\ 0 \neq x &\mapsto 1 \end{aligned}$$

has a solution over the Booleans [10]. Therefore, if the *support* of an empirical model is contextual, then so is the original model. The converse is not true, e.g. the $(2, 2, 2)$ Bell model from before has a non-contextual support [20]. This indicates that an empirical model whose support is contextual, such as the Hardy model, exhibits a stronger form of contextuality, which we call *logical contextuality*. Equivalently, a model is logically non-contextual if and only if, for every event $s \in \mathcal{O}^C$ in its support (i.e. $d_C(s) > 0$), there is a global assignment $t \in \mathcal{O}^{\mathcal{M}}$ such that $t|_C = s$. Since logical contextuality does not involve probabilities, it admits purely logical proofs, without the need for Bell inequalities [7].

De Silva translated the idea of logical contextuality to exclusivity graphs as follows [19].

Theorem 6.1. *An empirical model Ω is logically contextual if and only if the minimum independence degree over all vertices of the support graph G_Ω is less than the number of contexts.*

6.2 Strong contextuality

Now we take another look at the GHZ models from Section 3.3. It can be shown that for $n \geq 3$, these models exhibit an even stronger form of contextuality, which we call *strong contextuality*. A model is strongly contextual if, given any global assignment $t \in \mathcal{O}^{\mathcal{M}}$, there is some context C such that $t|_C$ is not in the support of the model (i.e. $d_C(t|_C) = 0$) [10]. Equivalently, a model is strongly contextual if and only if the propositional formulas defining its support, i.e. the formulas

$$\varphi_C = \bigvee_{\substack{s \in \mathcal{O}^C, \\ d_C(s) > 0}} \varphi_s$$

over all contexts C, are not simultaneously satisfiable [12].

If we compare this notion with that of standard contextuality, we see that this is much stronger, since it says that any global assignment is not even consistent with the support of the model, let alone the distributions. We also see that strong contextuality implies logical contextuality. However, the Hardy model is not strongly contextual, as it has the global assignment $\{a \mapsto 1, a' \mapsto 0, b \mapsto 1, b' \mapsto 0\}$ [10]. We thus have a strict hierarchy of strengths of contextuality:

contextual $<$ logically contextual $<$ strongly contextual.

AB used a modular-arithmetic approach to prove that the GHZ models are strongly contextual for all $n \geq 3$. AH proved that a model is strongly contextual if and only if it *maximally violates* a logical Bell inequality, which means that there is a logical Bell inequality with formulas φ_i such that $p(\varphi_i) = 1$ for all i. They then proved the strong contextuality of GHZ models where $n \geq 3$ by finding formulas that maximally violate a logical Bell inequality.

It is worth pointing out that strong contextuality is equivalent to the notion of *maximal contextuality*. Very roughly, this is the property that the model cannot be written as a convex sum of a non-contextual part and a non-signalling part where the coefficient of the non-contextual part is non-zero [10]. Thus a model is strongly/maximally contextual if and only if it can be written as a convex sum of contextual vertices of the non-signalling polytope over the cover $\mathcal{C}$ of the underlying measurement scenario [19, 21, 22].

De Silva proved an analogous result relating strong contextuality to exclusivity graphs.

Theorem 6.2. *An empirical model Ω is strongly contextual if and only if the independence number of its support graph G_Ω is less than the number of contexts.*

Note that, in ascending from logical contextuality to strong contextuality, we go from looking at the *minimum* independence degree, and comparing it with the number of contexts, to looking at the *maximum* independence degree.

7 Applications and future work

It is remarkable that we can equivalently view contextuality in seemingly very different ways. Furthermore, such myriad connections with other mathematical concepts allow us to borrow techniques from these areas. Sheaf theory is a powerful tool used throughout math to formalize local-to-global connections [10, 23]. Presheaves and global sections have also been studied in the context of the semantics of computation [24] and concurrency [25]. Abramsky, Mansfield and Barbosa studied contextuality using abelian presheaves and Čech cohomology [26]; AB mention that research of this nature is fairly new and there remains much to be discovered. Logical Bell inequalities involve expressions that have a similar form to basic weight formulas, which are used in artificial intelligence [27, 28, 12]. Furthermore, graph-theoretic methods have long had applications in quantum theory [21, 29] and a wealth of theory is available to us. Finally, there are formalisms for contextuality that we have not even touched on here, such as the work done by Spekkens et al. [30]. It would be useful to continue comparing these works with other formalisms thus far developed for contextuality, in a similar spirit to how equivalences and unifications have been described in this essay.

Research on contextuality as a resource for quantum advantage is very active. Significant progress has been made, but plenty of open problems remain. Howard et al. [31] proved that, for a single *qudit* of odd prime power dimension, contextuality with respect to stabilizer measurements is necessary for states that are suitable for magic state distillation. Their method uses Wigner functions and hence is not easily adaptable to the qubit case. De Silva commented that stronger forms of contextuality may be a useful property for identifying higher-qudit magic states [19]. For qubits, de Silva proved that, whenever $n > 1$, all n-qubit states are strongly contextual with respect to stabilizer measurements [19], extending the earlier Peres–Mermin argument for the case $n = 2$ [32]. This shows that even strong contextuality is not enough for identifying qubit magic states. However, an interesting direction could be to investigate whether it is possible to devise more specific measures that combine strong contextuality with other, scenario-specific properties. In more recent work, for the special case of strong *non-locality*, Abramsky et al. [33] found a family of three-qubit models exhibiting strong non-locality that use states from the GHZ SLOCC (stochastic local operations, classical communication) class. Many open problems arise from this work, such as finding an exhaustive collection of strongly non-local three-qubit models, investigating larger numbers of qubits, and understanding what these results say about resources for quantum speed-up.

Raussendorf [34] proved that a measurement-based quantum computer that computes a non-linear Boolean function with high probability is contextual, which among other implications builds a connection between contextuality and the quantum discrete logarithm algorithm, which is of significant practical interest. It would be interesting to find further connections between contextuality and other types of quantum computation, such as Pauli-based computation [35] or topological quantum computation [36].

8 Conclusion

We have seen elegant equivalences between logical, sheaf-theoretic and graph-theoretic formalisms for contextuality, standardizing the notation to allow us to seamlessly move between these frameworks. We have described three well-defined strengths of contextuality, and we have also discussed the many exciting avenues of investigation that can be followed in order to better understand contextuality (or non-locality) as a resource for quantum advantage. In the introduction, we mentioned that properties such as entanglement, superposition and discord have also been studied as potential resources for quantum speed-up in specific scenarios. To further develop the theory with a view to more applications, it would be interesting to try to marry these various concepts and see if it is possible to define more specific strengths of contextuality in this way.

References

[1] David Deutsch. "Quantum theory, the Church–Turing principle and the universal quantum computer". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 400.1818 (1985), pp. 97–117.

[2] Guifré Vidal. "Efficient classical simulation of slightly entangled quantum computations". In: *Physical review letters* 91.14 (2003), p. 147902.

[3] Maarten Van den Nest. "Universal quantum computation with little entanglement". In: *Physical review letters* 110.6 (2013), p. 060504.

[4] Ryszard Horodecki et al. "Quantum entanglement". In: *Rev. Mod. Phys.* 81 (2 June 2009), pp. 865–942. DOI: 10.1103/RevModPhys.81.865. URL: https://link.aps.org/doi/10.1103/RevModPhys.81.865.

[5] Animesh Datta, Anil Shaji, and Carlton M Caves. "Quantum discord and the power of one qubit". In: *Physical review letters* 100.5 (2008), p. 050502.

[6] Albert Einstein, Boris Podolsky, and Nathan Rosen. "Can quantum-mechanical description of physical reality be considered complete?" In: *Physical review* 47.10 (1935), p. 777.

[7] J. S. Bell. "On the Einstein Podolsky Rosen paradox". In: *Physics Physique Fizika* 1 (3 Nov. 1964), pp. 195–200. DOI: 10.1103/PhysicsPhysiqueFizika.1.195. URL: https://link.aps.org/doi/10.1103/PhysicsPhysiqueFizika.1.195.

[8] John F Clauser et al. "Proposed experiment to test local hidden-variable theories". In: *Physical review letters* 23.15 (1969), p. 880.

[9] Simon Kochen and EP Specker. "The Problem of Hidden Variables in Quantum Mechanics". In: *Journal of Mathematics and Mechanics* 17.1 (1967), pp. 59–87.

[10] Samson Abramsky and Adam Brandenburger. "The sheaf-theoretic structure of non-locality and contextuality". In: *New Journal of Physics* 13.11 (Nov. 2011), p. 113036. DOI: 10.1088/1367-2630/13/11/113036. URL: https://dx.doi.org/10.1088/1367-2630/13/11/113036.

[11] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2011. ISBN: 9781107002173.

[12] Samson Abramsky and Lucien Hardy. "Logical Bell inequalities". In: *Physical Review A* 85.6 (June 2012). ISSN: 1094-1622. DOI: 10.1103/physreva.85.062114. URL: http://dx.doi.org/10.1103/PhysRevA.85.062114.

[13] Lucien Hardy. "Nonlocality for two particles without inequalities for almost all entangled states". In: *Phys. Rev. Lett.* 71 (11 Sept. 1993), pp. 1665–1668. DOI: 10.1103/PhysRevLett.71.1665. URL: https://link.aps.org/doi/10.1103/PhysRevLett.71.1665.

[14] Daniel M Greenberger, Michael A Horne, and Anton Zeilinger. "Going beyond Bell's theorem". In: *Bell's theorem, quantum theory and conceptions of the universe*. Springer, 1989, pp. 69–72.

[15] Daniel M. Greenberger et al. "Bell's theorem without inequalities". In: *American Journal of Physics* 58.12 (Dec. 1990), pp. 1131–1143. ISSN: 0002-9505. DOI: 10.1119/1.16243. eprint: https://pubs.aip.org/aapt/ajp/article-pdf/58/12/1131/11479397/1131_1_online.pdf. URL: https://doi.org/10.1119/1.16243.

[16] Adán Cabello, Simone Severini, and Andreas Winter. "Graph-Theoretic Approach to Quantum Correlations". In: *Physical Review Letters* 112.4 (Jan. 2014). ISSN: 1079-7114. DOI: 10.1103/physrevlett.112.040401. URL: http://dx.doi.org/10.1103/PhysRevLett.112.040401.

[17] Sandu Popescu and Daniel Rohrlich. "Quantum nonlocality as an axiom". In: *Foundations of Physics* 24.3 (1994), pp. 379–385.

[18] Richard P. Feynman. "Negative Probability". In: *Quantum Implications: Essays in Honour of David Bohm*. Ed. by Basil J. Hiley and D. Peat. Methuen, 1987, pp. 235–248.

[19] Nadish de Silva. "Graph-theoretic strengths of contextuality". In: *Phys. Rev. A* 95 (3 Mar. 2017), p. 032108. DOI: 10.1103/PhysRevA.95.032108. URL: https://link.aps.org/doi/10.1103/PhysRevA.95.032108.

[20] Shane Mansfield and Tobias Fritz. "Hardy's Non-locality Paradox and Possibilistic Conditions for Non-locality". In: *Foundations of Physics* 42.5 (Mar. 2012), pp. 709–719. ISSN: 1572-9516. DOI: 10.1007/s10701-012-9640-1. URL: http://dx.doi.org/10.1007/s10701-012-9640-1.

[21] Antonio Acín et al. "A Combinatorial Approach to Nonlocality and Contextuality". In: *Communications in Mathematical Physics* 334.2 (Jan. 2015), pp. 533–628. ISSN: 1432-0916. DOI: 10.1007/s00220-014-2260-1. URL: http://dx.doi.org/10.1007/s00220-014-2260-1.

[22] Samson Abramsky et al. "Possibilities determine the combinatorial structure of probability polytopes". In: *Journal of Mathematical Psychology* 74 (2016). Foundations of Probability Theory in Psychology and Beyond, pp. 58–65. ISSN: 0022-2496. DOI: https://doi.org/10.1016/j.jmp.2016.03.006. URL: https://www.sciencedirect.com/science/article/pii/S0022249616300013.

[23] Saunders MacLane and Ieke Moerdijk. *Sheaves in geometry and logic: A first introduction to topos theory*. Springer Science & Business Media, 2012.

[24] Frank J Oles. "Type algebras, functor categories, and block structure". In: *DAIMI Report Series* 156 (1983).

[25] Gian Luca Cattani and Glynn Winskel. "Presheaf models for concurrency". In: *International Workshop on Computer Science Logic*. Springer. 1996, pp. 58–75.

[26] Samson Abramsky, Shane Mansfield, and Rui Soares Barbosa. "The Cohomology of Non-Locality and Contextuality". In: *Electronic Proceedings in Theoretical Computer Science* 95 (Oct. 2012), pp. 1–14. ISSN: 2075-2180. DOI: 10.4204/eptcs.95.1. URL: http://dx.doi.org/10.4204/EPTCS.95.1.

[27] Ronald Fagin, Joseph Y Halpern, and Nimrod Megiddo. "A logic for reasoning about probabilities". In: *Information and computation* 87.1-2 (1990), pp. 78–128.

[28] Nils J Nilsson. "Probabilistic logic". In: *Artificial intelligence* 28.1 (1986), pp. 71–87.

[29] Otfried Gühne et al. "Bell Inequalities for Graph States". In: *Phys. Rev. Lett.* 95 (12 Sept. 2005), p. 120405. DOI: 10.1103/PhysRevLett.95.120405. URL: https://link.aps.org/doi/10.1103/PhysRevLett.95.120405.

[30] Yeong-Cherng Liang, Robert W Spekkens, and Howard M Wiseman. "Specker's parable of the over-protective seer: A road to contextuality, nonlocality and complementarity". In: *Physics Reports* 506.1-2 (2011), pp. 1–39.

[31] Mark Howard et al. "Contextuality supplies the 'magic' for quantum computation". In: *Nature* 510.7505 (2014), pp. 351–355.

[32] N David Mermin. "Hidden variables and the two theorems of John Bell". In: *Reviews of Modern Physics* 65.3 (1993), p. 803.

[33] Samson Abramsky et al. "Minimum Quantum Resources for Strong Non-Locality". In: *12th Conference on the Theory of Quantum Computation, Communication, and Cryptography*. 2018.

[34] Robert Raussendorf. "Contextuality in measurement-based quantum computation". In: *Physical Review A* 88.2 (2013), p. 022322.

[35] Sergey Bravyi, Graeme Smith, and John A Smolin. "Trading classical and quantum computational resources". In: *Physical Review X* 6.2 (2016), p. 021043.

[36] A.Yu. Kitaev. "Fault-tolerant quantum computation by anyons". In: *Annals of Physics* 303.1 (Jan. 2003), pp. 2–30. ISSN: 0003-4916. DOI: 10.1016/s0003-4916(02)00018-0. URL: http://dx.doi.org/10.1016/S0003-4916(02)00018-0.