OPEN ACCESS
IOP Publishing
European Journal of Physics
Eur. J. Phys. **43** (2022) 055401 (22pp)
https://doi.org/10.1088/1361-6404/ac79e0

A demonstration of contextuality using quantum computers

Amine Laghaout, Altay Dikme, Nicolas Reichel and Gunnar Björk*

Department of Applied Physics, Royal Institute of Technology (KTH), Albanova University Center, 10691 Stockholm, Sweden

E-mail: gbjork@kth.se

Received 21 October 2021, revised 28 April 2022
Accepted for publication 17 June 2022
Published 7 July 2022

Abstract

Open-access, online quantum computers have shown significant improvements in the past decade. Although they still suffer from noise and scalability limitations, they do offer the possibility of experimenting with quantum circuits which would otherwise have required laboratory resources and prowesses beyond the reach of most students (and even researchers). In view of this, we revisit from the ground up the notion of contextuality and show that it can now be easily demonstrated on one of the IBM quantum computers. We showcase this with an implementation of the Peres–Mermin square which, despite the high error rates, manages to violate noncontextuality by almost 28 standard deviations.

Keywords: quantum contextuality, quantum computers, quantum non-demolition measurements, quantum realism

(Some figures may appear in colour only in the online journal)

1. Introduction

One of the very first principles taught in undergraduate quantum mechanics is the uncertainty principle. One learns, for instance, that a particle's position $\hat{x}$ cannot be measured together with its momentum $\hat{p}$. Mathematically, the principle is rather straightforward to explain, particularly in the wave function representation, where complementary variables can be visualized as Fourier transforms of one another. The more one variable is well-defined, the more is its transform spread out, and hence ill-defined, in the conjugate domain. However, physically—or

*Author to whom any correspondence should be addressed.

0143-0807/22/055401+22$33.00 Printed in the UK

rather, ontologically—the uncertainty principle is much more perplexing. All too often, it is explained away by invoking a (presumed) disturbance caused by the measurement process whereby, say, the retrieval of x collapses the wave function, and therefore irrevocably corrupts any pre-existing information that it could have contained about p. This widespread explanation, which perfectly abides by the orthodoxy of the Copenhagen interpretation, unfortunately obscures a more fundamental question. In particular, it does not say whether there actually *was* any pre-existing value of p to be disturbed in the first place. If anything, it tacitly suggests such pre-existence of a well-defined p since, after all, one could have chosen to measure it first instead of x. One can thus naively conclude that uncertainty is wholly due to disturbances, i.e., to complex but nonetheless deterministic processes which are governed by hidden variables. Such variables, which act both at the state-preparation and at state-measurement stages, just happen to be unfathomable with our current level of technology (and perhaps intelligence). The question of whether hidden variables are indeed 'pulling the strings' behind a facade of stochasticity was raised as early as the 1930s, most notably in the Einstein–Podolsky–Rosen (EPR) paper [1]. It was then formalized in the 1960s by the works of Bell [2], and Kochen and Specker [3], and finally answered by a host of experiments in the following decades [4–17]. Notwithstanding the likelihood of some pedantically elaborate loopholes [18], most results so far concur in their discredit of a hidden variable theory [19].

The aforementioned proposals and experiments, though approaching the problem of hidden variables from different angles, revolve around one core question: they essentially all investigate whether physical systems have well-defined, pre-existing properties independently of measurement. (See figure 1 for an intuitive visualization of the problem.) In this paper, we attempt a walk-through of such an investigation based on a scheme known as the Peres–Mermin square [20–22]. We begin in section 2 by building a basic intuition of what it means for a system to have pre-existing properties. The notion of contextuality is then elicited as arising from scenarios where pre-existing properties cannot logically account for the measurement outcomes. We first illustrate this with a toy model which we then transpose in section 3 to the quantum domain with the Peres–Mermin square. Finally, section 4 describes our implementation of the Peres–Mermin square on one of the IBM quantum computers. Our results, though degraded by noise, succeed nonetheless in demonstrating with high confidence that the Peres–Mermin square cannot be described by pre-existing properties and instead responds contextually to measurement choices.

All in all, the pedagogical aim of our presentation lies in the conceptual compactness of the Peres–Mermin square together with the open-access and ease-of-use of quantum computers. The latter are indeed becoming a versatile tool for experimentation, both for educational and for research purposes [23–26].

2. Background

2.1. Classical systems as ensembles of possibilities

A central premise of classical physics is that any system $\mathcal{S}$ is characterized by a tuple

$$\vec{a} = [a_1, \ldots, a_n, \ldots, a_N]^\mathsf{T} \tag{1}$$

of N attribute properties, e.g., its position, momentum, mass, spin, etc. Classical measurements on the system are considered to be deterministic functions onto real scalars,

$$\hat{a}_n : \mathcal{S} \to a_n \in \mathbb{R}, \tag{2}$$

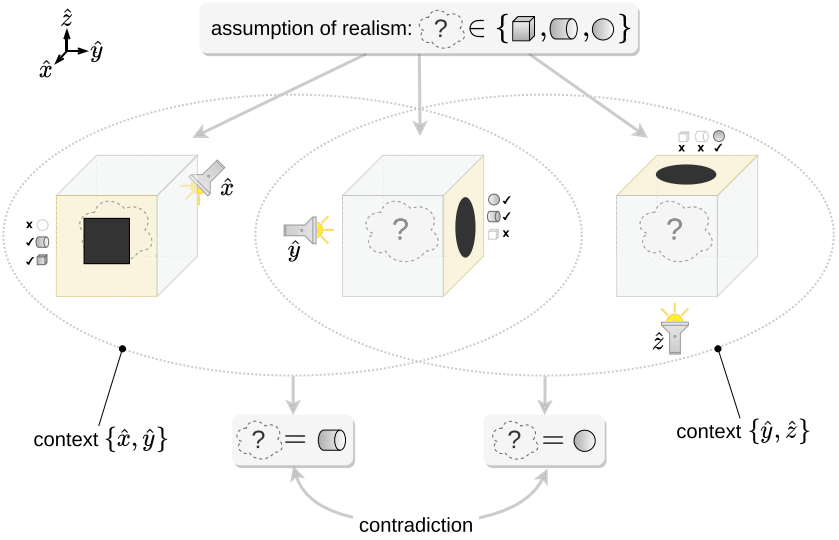

Figure 1. Illustration of contextuality in three dimensions. A mystery object is assumed to be pre-existing, i.e., constitutes an 'element of reality', if it is drawn with unbiased replacement from an ensemble of objects with well-defined properties. If so, over several runs of the experiment, its measurement outcomes (i.e., the shadows it casts on a screen) should be consistent independently of the context (i.e., the angles of illumination). In the present example, the mystery object is illuminated from pairs of directions, namely from behind ($\hat{x}$) and from the left ($\hat{y}$) in the context $\{\hat{x}, \hat{y}\}$, and from the left ($\hat{y}$) and from the bottom ($\hat{z}$) in context $\{\hat{y}, \hat{z}\}$. Whereas one is lead to conclude that the object is the cylinder in the former context, a copy of that same object is instead inferred to be the sphere in the latter context. This in turn contradicts the initial condition for realism. Note that we are not illuminating the object in all three Cartesian dimensions at once so as to play by the rules of the uncertainty principle which, in this case, is reflected by the fact that the object cannot be jointly illuminated from behind and from the bottom. It can only be shone upon from specific combinations of angles—namely those that would correspond to quantum-mechanically commuting observables.

which merely reveal the pre-existing values of the attributes without disrupting them. For example, if some system is in a state $\vec{a} = [a_1 = +1, a_2 = -1, a_3 = -1]^\intercal$, then any measurement $\hat{a}_2$ of a_2 will consistently return -1 regardless of whether a_1 or a_3 were also measured. If this were not the case, or if successive measurements yielded different results, then classical physics can still account for such an apparent indeterminism by modeling the system as a statistical ensemble from which a particular instance is randomly sampled. This way, one does not renounce the assumption that the tuple $\vec{a}$ of attributes is pre-existing and independent of measurement, but simply encapsulates the observer's epistemic ignorance about the ensemble with a probabilistic description. Classically, the most general description of the state of a system is therefore that of a statistical mixture, namely an ensemble which is a convex combination of all the K possible attribute tuples

$$\langle \vec{a} \rangle = \left[\vec{a}^{(1)}, \ldots, \vec{a}^{(k)}, \ldots, \vec{a}^{(K)} \right] \vec{p} = \hat{A} \vec{p}, \tag{3}$$

where the kth element of the vector $\vec{p}$ weighs the likelihood that the kth possible tuple $\vec{a}^{(k)}$ of pre-existing attributes is picked out from the ensemble. Here, $\hat{A}$ is the assemblage of all

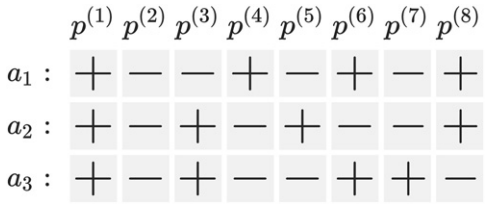

Figure 2. Each column is an ontic state of a three-attribute, binary system. Together, they make up the ensemble of possible values that the attributes of the system can be in. Since we have three binary attributes $a_i \in \{-1, +1\}$, the cardinality of this ensemble is $2^3 = 8$. A general classical description of the system is given by the joint probability mass function $\vec{p}$ which weighs the different ontic states.

possible tuples into an $N \times K$ matrix. The convexity of equation (3) implies that $\vec{p}$ is a probability distribution, i.e., that

$$p^{(k)} \in [0, 1], \quad \forall k \in \{1, \ldots, K\} \quad \text{and} \quad \sum_{k=1}^{K} p^{(k)} = 1. \tag{4}$$

Note that unlike $\vec{a}$ in equation (1), $\langle \vec{a} \rangle$ in equation (3) does not necessarily represent the actual state upon any given measurement, but is simply the expectation value over such possible states.

We shall later see how the above interpretation of measurements as random samplings from a static and pre-existing grab bag is invalidated by quantum mechanics. For now, let us illustrate it with a simple example which shall set the stage for our later discussion of contextuality. Let us consider some system which is presumably fully determined by three binary attributes $\vec{a} = [a_1, a_2, a_3]^\intercal$ where $a_1, a_2, a_3 \in \{-1, +1\}$. Prior to measurement, the 'realm of possibilities' for this system—formally referred to as the set of ontic states [27]—consists of the $K = 2^3 = 8$ permutations which are displayed as columns in figure 2.

The characterization of a general system consists of narrowing down this realm of possibilities to the joint probability mass function $\vec{p}$ that weighs each of the columns. (Recall that, for the observer, $\vec{p}$ serves as the most complete description of the ensemble.) Assuming that one can sample many copies of the system at will, the most straightforward approach would be to repeatedly measure the full triplet of attributes and then normalize the occurrences of each combination $\vec{a}^{(k)}$ to obtain its corresponding probability $p^{(k)}$. This approach is not always transposable to quantum mechanics, however, since the non-commutativity of certain pairs of attributes would bar us from measuring them together. In order to allow for an adaptation to the quantum case, we shall therefore constrain ourselves to measure only one attribute at a time such that, over several runs, we can build up the expectation values $\langle a_1 \rangle$, $\langle a_2 \rangle$, and $\langle a_3 \rangle$ of the individual attributes. If one assumes that the expectation value $\langle \vec{a} \rangle$ of the tuple is equal to the tuple of expectation values—as would be reasonable if the ensemble distribution is independent on the choice of measurement—then all that remains is to solve

$$\begin{bmatrix} \langle a_1 \rangle \\ \langle a_2 \rangle \\ \langle a_3 \rangle \end{bmatrix} = \begin{bmatrix} +1 & -1 & -1 & +1 & -1 & +1 & -1 & +1 \\ +1 & -1 & +1 & -1 & +1 & -1 & -1 & +1 \\ +1 & -1 & +1 & -1 & -1 & +1 & +1 & -1 \end{bmatrix} \vec{p} \tag{5}$$

for $\vec{p}$ provided that equation (4) is satisfied.

2.2. Contexts as sets of no-disturbance

2.2.1. Sets of compatible measurements. In the discussion leading up to equation (5), we steered clear of issues that could arise from the uncertainty principle by measuring only one attribute for each copy of the system. This restriction can however be relaxed whenever quantum mechanics allows it, namely whenever the quantum mechanical observables that correspond to the measured attributes commute. For example, if the quantum mechanical operator corresponding to measurement $\hat{a}_n$ commutes with that of $\hat{a}_m$, then the pre-existing values of both a_n and a_m can be faithfully retrieved from the *same* sample of the system ensemble without having to worry about measurement-induced disturbances. This can be extended to include any other compatible measurements, thereby forming a set

$$\mathcal{M} = \{\hat{a}_n, \hat{a}_k, \hat{a}_l, \hat{a}_m, \ldots\} \tag{6}$$

which is referred to as a *context*. Any sequence of measurements picked out from the same context can be performed in any order and is expected to return consistent statistics. If there exists a different context

$$\mathcal{M}' = \{\hat{a}_n, \hat{a}_\kappa, \hat{a}_\lambda, \hat{a}_\nu, \ldots\}, \tag{7}$$

where one of the elements is common to both contexts is, say,

$$\hat{a}_n \in \mathcal{M} \cap \mathcal{M}', \tag{8}$$

then *noncontextuality* is the assertion that the statistics[1] inferred about a_n are independent of which context it was measured in. (No such guarantee of determinism can be made for the remaining attributes $a_m, a_k, a_\nu, a_\kappa, \ldots$, however, unless they too are common to both contexts.) In contrast, *contextuality* arises whenever measurement statistics do depend on the context as shall be shown in section 2.4.

2.2.2. Product measurements. Measurements are not limited to gathering the expectation values $\langle a_n \rangle$ of the individual attributes. As we just saw, they can also consist of sequences—or, more specifically, products—of the individual attributes, so long as those attributes belong to the same context. An example could be

$$\hat{a}_{(n,\nu)} : \mathcal{S} \to a_n a_\nu \in \mathbb{R}, \tag{9}$$

provided that the corresponding quantum observables commute, i.e., $[\hat{a}_n, \hat{a}_\nu] = 0$. (Note that the order of the subscript is irrelevant thanks to the commutativity property, i.e., $\hat{a}_{(n,\nu)} = \hat{a}_{(\nu,n)}$.) These sequence measurements, just like the attributes that make them up, can in turn be grouped into contexts of mutually compatible operators, e.g.,

$$\mathcal{M}'' = \{\hat{a}_{(n,k)}, \hat{a}_{(\nu,n)}, \ldots\}. \tag{10}$$

Note a crucial caveat: although the joint measurements in the context are compatible, the individual measurements they are composed of may not be. E.g., $[\hat{a}_{(n,k)}, \hat{a}_{(\nu,n)}] = 0$ in equation (10) does not guarantee that $[\hat{a}_k, \hat{a}_\nu] = 0$.

[1] Strictly speaking, non-contextuality implies that there should exist at least one probability distribution that is consistent with those statistics across *all* contexts.

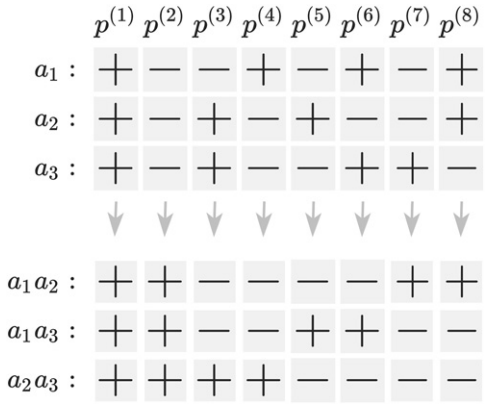

Figure 3. Many-to-one relationship between ontic states and the vector of joint product measurements $a_i a_j$. While the former live in an ensemble of $2^3 = 8$ possibilities, the latter only span half as many combinations. The remaining combinations—e.g., $[-1, +1 + 1]^\intercal$—are impossible to obtain from the original, predefined ensemble of ontic states.

2.3. Possible vs 'impossible' measurement outcomes

Let us extend our earlier example of a three-attribute system with the pairwise product measurements we just introduced in equation (9). As can be seen from figure 3, although we have $a_n a_m \in \{-1, +1\}$ for each individual product, the joint set of triplets $[a_1a_2, a_1a_3, a_2a_3]$ does not cover all the eight possible combinations in $\{-1, +1\}^3$ due to the degeneracy of multiplication, namely $(\pm 1) \times (\pm 1) = +1$ and $(\pm 1) \times (\mp 1) = -1$. In terms of product measurements, the realm of possibilities has effectively been halved in comparison to those of equation (5) as we now have

$$\underbrace{\begin{bmatrix}\langle a_1a_2\rangle\\ \langle a_1a_3\rangle\\ \langle a_2a_3\rangle\end{bmatrix}}_{\vec{\mu}} = \underbrace{\begin{bmatrix}+1 & +1 & -1 & -1 & -1 & -1 & +1 & +1\\ +1 & +1 & +1 & +1 & -1 & -1 & -1 & -1\\ +1 & +1 & -1 & -1 & +1 & +1 & -1 & -1\end{bmatrix}}_{\hat{A}} \vec{p}, \tag{11}$$

which, due to the column-degeneracy of $\hat{A}$, can be rewritten in more compact form as

$$\vec{\mu} = \underbrace{\begin{bmatrix}+1 & -1 & -1 & +1\\ +1 & +1 & -1 & -1\\ +1 & -1 & +1 & -1\end{bmatrix}}_{\hat{A}^{(\mathrm{R})}} \vec{p}_\pi, \tag{12}$$

with $\vec{p}_\pi$ now being four-dimensional[2]. A consequence of this reduction in the realm of possibilities is that certain measurement vectors such as

$$\vec{\mu} = [-1, +1, +1]^\intercal \tag{13}$$

[2] Note that equation (11) only makes sense if the attributes a_n are probed individually. If the measurement merely returns the joint product, then only equation (12) makes informational sense to the observer.

become impossible to obtain since they cannot solve equation (12) under the constraint of convexity. In other words, no fixed, pre-existing ensemble of ontic states can ever lead to such an outcome $\vec{\mu}$. One can see that the distinction between classically possible and impossible measurement outcomes has an algebraic underpinning: the convex hull spanned by the complete set $\{-1,+1\}^3$ of combinations is broken down into two convex hulls. The first is fully consistent with classical assumptions (figure 3) and is spanned by the polytope delimited by the columns of

$$\hat{A}^{(\mathrm{R})} = \begin{bmatrix} +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \end{bmatrix}. \tag{14}$$

Any measurement vector $\vec{\mu}$ on the surface of this hull is deemed *real* as per the EPR definition of realism. The second hull is delimited by the remaining set of combinations

$$\hat{A}^{(\mathrm{U})} = \begin{bmatrix} -1 & +1 & +1 & -1 \\ -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 \end{bmatrix}. \tag{15}$$

We shall refer to any measurement result that lies on this latter hull as *unreal*. (A non-mathematical walk-through of convex hulls and their relevance as presented above is provided in appendix A.)

In light of all the above, we are able to define a metric for violations of realism. Indeed, if some measurement vector fails to satisfy the convex equation $\vec{\mu} = \hat{A}^{(\mathrm{R})}\vec{p}_\pi$, then its Euclidean distance

$$\mathcal{D}^{(\mathrm{R})}_{\vec{\mu}} = \min_{\vec{p}_\pi} \|\vec{\mu} - \hat{A}^{(\mathrm{R})}\vec{p}_\pi\| \tag{16}$$

from the real hull is a quantifier of its remoteness from realism.

2.4. Contextuality and unreal measurements

What if one *does* obtain an unreal—i.e., classically-impossible—measurement vector as in equation (13), either empirically or via the predictions of a non-classical theory such as quantum mechanics? Should one then abandon the assumption that the attributes are predefined? For example, could it be that, unbeknownst to the experimenter, the joint probability distribution $\vec{p}$ changes its profile in response to the choice of context? In order to visualize these interpretations, figure 4 renders our three-attribute system as a graph where each vertex represents a binary attribute. The different contexts $\mathcal{M}_{ij} = \{\hat{a}_i, \hat{a}_j\}$ are highlighted as the sets of connected vertices (i.e., hyperedges) that can be queried simultaneously[3]. For each context, figure 5 then highlights the joint probability distributions $\vec{p}$ that would be consistent with a real and unreal measurement vector, respectively. An immediate observation is that, for a real vector such as $[-1,+1,-1]^\intercal$, one finds ontic combinations which are consistent across all contexts, namely those weighted by $p^{(5)}$ and $p^{(6)}$ in this example. For the unreal vector $[-1,+1,+1]^\intercal$, however, no such ontic combination can be found which would be consistent across all contexts. In other words, it is as if the unreal measurement vector only makes sense when examined

[3] In this trivial example, all attributes commute with one another, so there is in reality one big context covering them all. The current grouping of the contexts as pairs of attributes is therefore contrived, but nonetheless valid for illustration purposes. The Peres–Mermin square presents a more realistic case where the different attributes are indeed required to belong to separate contexts because of their incompatibility.

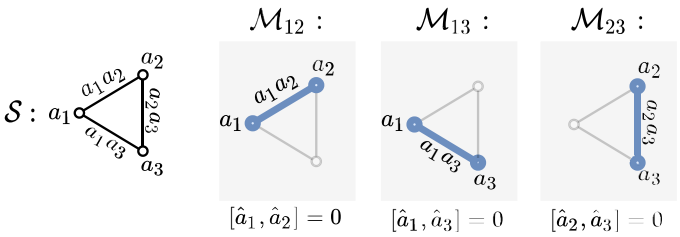

Figure 4. If one disposes the attributes a_i of a system $\mathcal{S}$ as the vertices of a graph, then contexts are formed by the hyperedges covering mutually compatible vertices. Here, the tripartite binary-attribute system we have discussed so far is arbitrarily broken down into three contexts where each context is formed by a pair of attributes. For quantum-mechanical systems in general, the contexts are dictated by the commutation rules between the attributes.

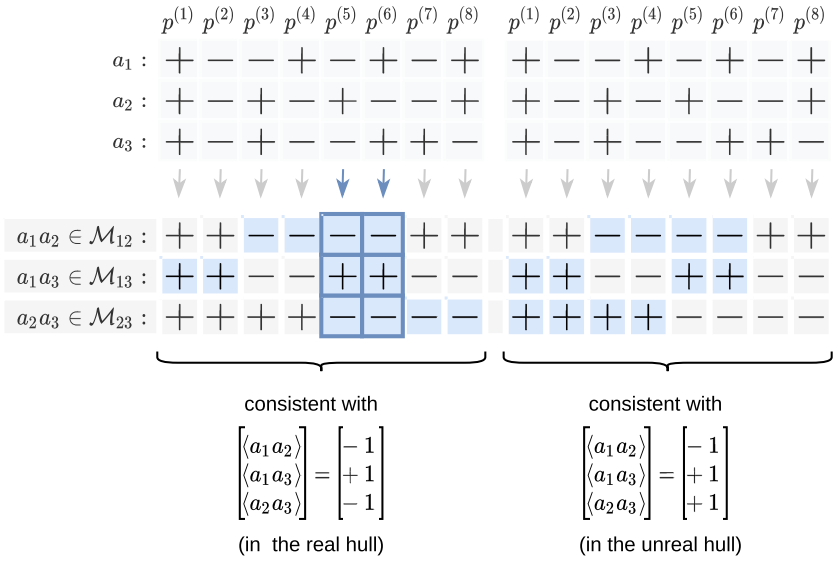

Figure 5. Illustration of contextuality using a toy model of three binary attributes. All eight ontic combinations of these attributes are displayed at the top alongside their associated weights $\vec{p}$ in the presumed ensemble distribution. For each of the contexts defined in figure 4, the measurement outcomes that match the examples of real and unreal vectors are highlighted in blue in the left and right columns, respectively. Consider first the vector $[-1, +1, -1]^\intercal$ in the real hull. All three contexts have some non-zero overlap, namely at $p^{(5)}$ and $p^{(6)}$. In other words, for this particular real measurement vector, the underlying joint probability distribution reduces to zero everywhere except for $p^{(5)}$ and $p^{(6)}$. In contrast, the unreal vector $[-1, +1, +1]^\intercal$ does not have global consistency since the inferred ontic combinations seem to depend on the contexts which are used for probing the joint probability distribution. E.g., the latter collapses onto $p^{(5)}$ and $p^{(6)}$ when probed in the contexts $\mathcal{M}_{12}$ and $\mathcal{M}_{13}$ but onto $p^{(1)}$ and $p^{(2)}$ when probed in the contexts $\mathcal{M}_{13}$ and $\mathcal{M}_{23}$.

in *some* contexts, but ends up contradicting itself when examined in *all* contexts—hence the term contextuality. A discussion of whether contextuality and violations of realism are just facets of the same phenomenon or two fundamentally distinct ones is beyond the scope of this article. What is nonetheless evident is that they both challenge the classical understanding of state preparation $\vec{p}$ as a cause and state measurement $\langle\vec{a}\rangle$ as its effect [28, 29]. For our purposes, we shall content ourselves with Abramsky's [30] minimalist definition of contextuality as a scenario where measurements statistics are 'locally consistent but globally inconsistent'. Here, locality does not necessarily refer to the spatial dimension but, more generally, to the graph topology [31] formed by the contexts: measurements within a given context or between a subset of contexts are considered local, whereas those that span all contexts are considered global[4].

A remark is in order regarding the role of non-commutativity: we presented it as a restriction on the measurements that could be performed together, and thereby as the rationale for grouping measurements into contexts. Indeed, if it were not for non-commutativity, there would not be any contextuality since there would not be any contexts to speak of. Consequently, one should regard the existence of incompatible observables as integral to the notion of contextuality, and not merely an auxiliary constraint on it.

3. The Peres–Mermin square

3.1. Motivation: state-independence

Our discussion of contextuality has so far only invoked algebra and logic. We shall now review a way to demonstrate it physically. It has been known for decades that quantum mechanics can lead to contextual statistics [3]. Work on non-locality [2], which can be regarded as a special case of contextuality where the sets of no-disturbance arise from the space-like separation of the measurements (as opposed to their commutativity), has been particularly prolific [32]. Experimental demonstrations of non-locality [19], and more explicitly of contextuality [4–17], have repeatedly confirmed that quantum mechanical systems behave contextually. These demonstrations are however demanding to perform due to their vulnerability to quantum decoherence both at the state-engineering and state-measurement stages. For the purposes of this article, we shall revisit a proposal which is state-independent and therefore does away with the specificities of state preparation altogether [33, 34]. This proposal, known as the Peres–Mermin square, was first realized in 2009 using an elaborate ion trap setup [6]. The following decade has since witnessed the democratization of quantum computing, thereby making it feasible to reproduce such demanding experiments using inexpensive and user-friendly interfaces to online quantum computers [23–26]. The process is as easy as writing a set of instructions to be uploaded to the remote computer, which subsequently logs the measurement outcomes to a downloadable file. It has thus become possible for virtually anyone with programming skills and a basic understanding of qubit operations to observe first-hand the counter-intuitive effects of contextuality. The Peres–Mermin proposal is particularly suited for this task since, as we shall see next, it only requires conceptually simple qubit measurements.

[4] Bell nonlocality is a special case of contextuality where space-like-separated measurement sites belong to the same context since, as per the no-signaling theorem, no disturbance of the system is to be (classically) expected between those sites.

3.2. The square

3.2.1. Contexts with deterministically unreal outcomes. In terms of state engineering, the Peres–Mermin setup requires nothing more than an arbitrary two-qubit system—be it pure or not. The classical assumption is that each qubit points in a pre-existing, albeit stochastic, direction of the Bloch sphere [35] and that a projective measurement along one of the three basis vectors merely collapses the qubit onto its x, y, or z components[5]. The operators behind these three projections, the Pauli operators,

$$\hat{\sigma}_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \hat{\sigma}_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \hat{\sigma}_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \tag{17}$$

do not to commute and therefore cannot be grouped into contexts. However, as one deals with a *two*-qubit system, it was ingeniously shown by Mermin [21], based on earlier work by Peres [20], that not only do certain pairs of *joint* projection operators commute, but that the contexts they thus form *always* produce the same unreal measurement vector regardless of the input state. This property is enabled by a fourth operator, namely the identity

$$\hat{\mathbb{I}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{18}$$

which leaves unchanged any qubit it operates on. (Physically, such a qubit would not be measured at all.) The six contexts thus constructed,

$$\mathcal{M}_{r_1} = \left\{\hat{\sigma}_x \otimes \hat{\mathbb{I}}, \hat{\mathbb{I}} \otimes \hat{\sigma}_x, \hat{\sigma}_x \otimes \hat{\sigma}_x\right\}, \tag{19}$$

$$\mathcal{M}_{r_2} = \left\{\hat{\mathbb{I}} \otimes \hat{\sigma}_y, \hat{\sigma}_y \otimes \hat{\mathbb{I}}, \hat{\sigma}_y \otimes \hat{\sigma}_y\right\}, \tag{20}$$

$$\mathcal{M}_{r_3} = \left\{\hat{\sigma}_x \otimes \hat{\sigma}_y, \hat{\sigma}_y \otimes \hat{\sigma}_x, \hat{\sigma}_z \otimes \hat{\sigma}_z\right\}, \tag{21}$$

$$\mathcal{M}_{c_1} = \left\{\hat{\sigma}_x \otimes \hat{\mathbb{I}}, \hat{\mathbb{I}} \otimes \hat{\sigma}_y, \hat{\sigma}_x \otimes \hat{\sigma}_y\right\}, \tag{22}$$

$$\mathcal{M}_{c_2} = \left\{\hat{\mathbb{I}} \otimes \hat{\sigma}_x, \hat{\sigma}_y \otimes \hat{\mathbb{I}}, \hat{\sigma}_y \otimes \hat{\sigma}_x\right\}, \quad \text{and} \tag{23}$$

$$\mathcal{M}_{c_3} = \left\{\hat{\sigma}_x \otimes \hat{\sigma}_x, \hat{\sigma}_y \otimes \hat{\sigma}_y, \hat{\sigma}_z \otimes \hat{\sigma}_z\right\}, \tag{24}$$

can be arranged as the rows and columns of the square in table 1. As can be verified by simple matrix multiplication, the product of measurements drawn from any given context is tantamount to the identity operation, except for the last column, which yields the negative identity. (A full derivation for context $\mathcal{M}_{c_3}$ is provided in appendix B.) Quantum mechanically, therefore, the eigenvalues of these product measurements are deterministic, namely $+1$ for the rows and first two columns and -1 for the third column.

This leads to a logical paradox: if one were to accept the classical premise that the outcome of all nine joint measurements in the square are (i) predetermined to be either $+1$ or -1, and (ii) that they are independent of whether they are obtained in the context of a row or column sequence, then none of the $2^9 = 512$ possible assignments of the square—or convex

[5] Such a collapse can be stochastic as per the ensemble interpretation of the measurement outcomes.

Table 1. The Peres–Mermin square is an assemblage of joint qubit measurements which, when multiplied along the rows and columns yield definite eigenvalues, namely $+1$ for the rows and first two columns, and -1 for the last column.

	$\mathcal{M}_{c_1}$	$\mathcal{M}_{c_2}$	$\mathcal{M}_{c_3}$	
$\mathcal{M}_{r_1}$	$\hat{\sigma}_x \otimes \hat{\mathbb{I}}$	$\hat{\mathbb{I}} \otimes \hat{\sigma}_x$	$\hat{\sigma}_x \otimes \hat{\sigma}_x$	$+\hat{\mathbb{I}}$
$\mathcal{M}_{r_2}$	$\hat{\mathbb{I}} \otimes \hat{\sigma}_y$	$\hat{\sigma}_y \otimes \hat{\mathbb{I}}$	$\hat{\sigma}_y \otimes \hat{\sigma}_y$	$+\hat{\mathbb{I}}$
$\mathcal{M}_{r_3}$	$\hat{\sigma}_x \otimes \hat{\sigma}_y$	$\hat{\sigma}_y \otimes \hat{\sigma}_x$	$\hat{\sigma}_z \otimes \hat{\sigma}_z$	$+\hat{\mathbb{I}}$
	$+\hat{\mathbb{I}}$	$+\hat{\mathbb{I}}$	$-\hat{\mathbb{I}}$	

combinations thereof—can satisfy the predictions of quantum mechanics. Indeed, the row contexts generate eigenvalues of $+1$, thereby requiring an even number of -1 entries in each row and thus an even number of -1s in the square. In contrast, the first two columns generate eigenvalues of $+1$ while the last column yields -1, thus implying an overall odd number of -1s in the square. Clearly, the assumptions (i) and (ii) cannot hold simultaneously in quantum mechanics: either (i) is true and we lose noncontextuality, or (ii) is true and we lose realism—or we lose both together.

3.2.2. The unreal measurement vector. To better understand how the above paradox is a manifestation of contextuality, let us draw the parallel with our toy example so as to come up with the Peres–Mermin analog of the unreal measurement vector equation (13).

Denote the entries of the square as $\hat{a}_{ij}$, where i and j are the indices of the rows and columns, respectively, e.g., $\hat{a}_{21} = \hat{\mathbb{I}} \otimes \hat{\sigma}_y$. The values that these attributes can take on upon measurement are $a_{ij} \in \{-1, +1\}$ such that the nine-dimensional ontic states

$$\vec{a} = [a_{11}, a_{12}, a_{13}, a_{21}, a_{22}, a_{23}, a_{31}, a_{32}, a_{33}]^\intercal \tag{25}$$

are drawn from a set of $2^9 = 512$ combinations. Just as in the toy example, however, the distinction between real and unreal measurements does not arise from the 'base' ontic attributes themselves, but rather from logical combinations thereof, namely product measurements. For the Peres–Mermin system, these products can only be performed within its six contexts, i.e., from its three rows and three columns

$$\hat{r}_i = \prod_{j=1}^{3} \hat{a}_{ij} \quad \text{and} \quad \hat{c}_j = \prod_{i=1}^{3} \hat{a}_{ij}, \tag{26}$$

respectively (figure 6). The measurement vector we are interested in could therefore be the concatenation of the row and column outcomes,

$$\vec{\mu} = \vec{r} \oplus \vec{c} = \left[\langle r_1 \rangle, \langle r_2 \rangle, \langle r_3 \rangle, \langle c_1 \rangle, \langle c_2 \rangle, \langle c_3 \rangle\right]^\intercal. \tag{27}$$

Classically, although each row or column can *individually* yield a value in $\{-1, +1\}$, the complete six-dimensional measurement vector $\vec{\mu}$ does not cover all $2^6 = 64$ combinations. This is because, as per equation (26), the ontic states of equation (25) only map to 32 of those combinations. In particular, the prediction of quantum mechanics, whereby the context measurements have the deterministic eigenvalues

$$\vec{\mu}^{(\mathrm{Q})} = [+1, +1, +1, +1, +1, -1]^\intercal, \tag{28}$$

cannot be mapped by the ontic states. In other words, just like equation (13), these measurement outcomes cannot arise from a convex combination of the vectors supporting the real hull.

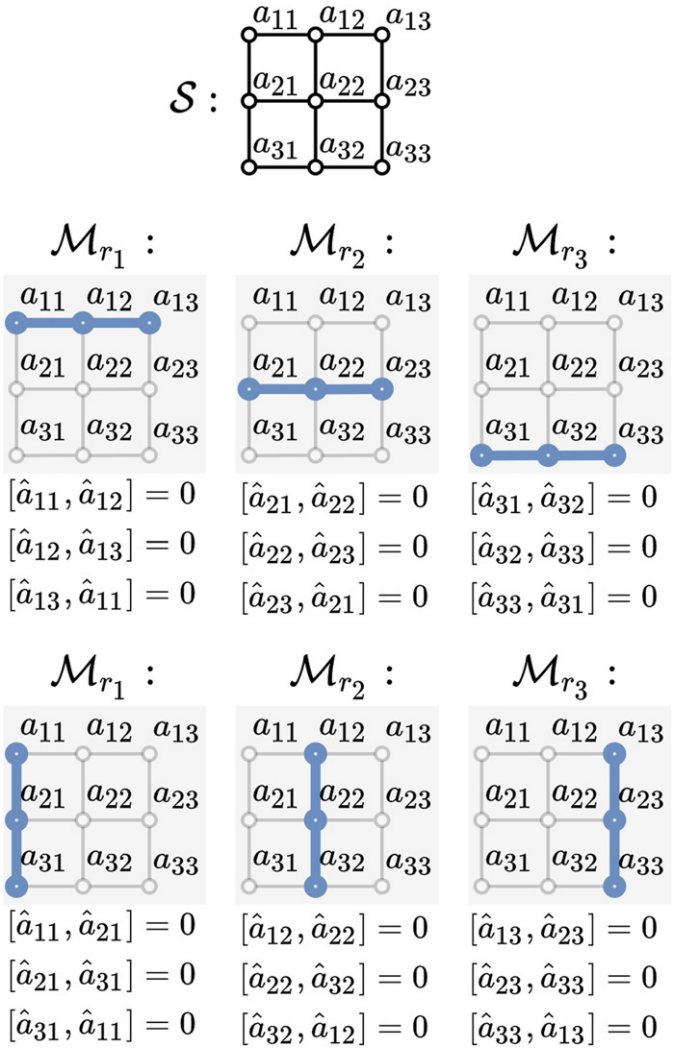

Figure 6. The six contexts of the Peres–Mermin square.

3.3. The real and unreal hulls

Let us now explicitly identify the contours of the real and unreal hulls that arise from the Peres–Mermin square. Realism implies that the attributes of the square have pre-existing values in $\{-1, +1\}$. Consequently, the row and column measurements should agree as to the overall parity of the square; i.e., they should agree as to whether there exists an even or and odd number of -1s. Hence, realism requires that $r_1 \cdot r_2 \cdot r_3 = c_1 \cdot c_2 \cdot c_3$. This condition can be re-packaged as the Boolean function

$$f(\vec{\mu}) = \frac{1}{2}\left(\prod_{i=1}^{6} \mu_i + 1\right), \tag{29}$$

which yields 1 (true) when $\vec{\mu}$ is in the real hull and 0 (false) when $\vec{\mu}$ is in the unreal hull. The 6×32 matrices $\hat{A}^{(\mathrm{R})}$ and $\hat{A}^{(\mathrm{U})}$ corresponding to equations (14) and (15), respectively are therefore made up of the combinations in $\{-1,+1\}^6$ that satisfy $f(\vec{\mu}) = 1$ and $f(\vec{\mu}) = 0$, respectively. In particular, a simple check with $f(\vec{\mu}^{(\mathrm{Q})}) = 0$ confirms that quantum mechanics produces an unreal vector. Even more telling is the distance of $\vec{\mu}^{(\mathrm{Q})}$ to the real hull, which can be shown to be

$$\mathcal{D}^{(\mathrm{R})}_{\vec{\mu}^{(\mathrm{Q})}} = \sqrt{\frac{2}{3}} \approx 0.8165, \tag{30}$$

using constrained nonlinear optimization on equation (16).

As we shall see in the next section, our goal is to check that the measurement vector $\vec{\mu}^{(\mathrm{IBM})}$ obtained experimentally on the quantum computer approximates the vector $\vec{\mu}^{(\mathrm{Q})}$ predicted by quantum mechanics. More specifically, we aim to verify that the distance $\mathcal{D}^{(\mathrm{R})}_{\vec{\mu}^{(\mathrm{IBM})}}$ is large enough to ascertain a violation of realism beyond statistical uncertainty.

4. Measuring the Peres–Mermin square on a quantum computer

4.1. Characterization of the quantum hardware

We ran our experiments on the IBM Quantum Experience platform, which is publicly available on the internet [36]. Through its online portal, we had access to three quantum computers, `ibmqx2` 'Yorktown', `ibmqx4` 'Tenerife', and `ibmqx16` 'Melbourne', which all fall under the category of Noisy Intermediate-Scale Quantum Computers (NISQs). For the present work, we ran our experiments on `ibmqx4`. It consists of five qubits interconnected in a butterfly pattern such that qubits 0, 1 and 2 are interconnected, and so are qubits 2, 3, and 4. Qubit 2 is therefore the only one that is fully connected—and hence can be measured jointly—with all the others.

The qubits are furthermore prone to decoherence, thereby requiring several runs of the experiment to make up for statistical errors. In particular, IBM's characterization of `ibmqx4` showed that the relaxation time T_1 for the different qubits lies between 30.1 μs and 52.1 μs, whereas the dephasing time T_2 lies between 4.9 μs and 53.0 μs. Similarly, the single- and two-qubit gate errors show a significant spread of 0.69% to 3.37% and of 2.12% to 7.95%, respectively. Finally, the readout errors on the qubit range from 3.4% to 34.8% (!) depending on the qubit.

Finally, as a preamble to measuring the full rows and columns of the Peres–Mermin square, we performed measurements of the nine individual observables on different input states on the `ibmqx4` computer. For example, the results when measuring the input state $|-1_z\rangle \otimes |-1_z\rangle$, where the notation -1_z means that it is the eigenstate to $\hat{\sigma}_z$ having eigenvalue -1, yielded the statistics in table 2 when repeated 8192 times. (This is the maximum number of repetitions IBM allows ordinary users on any given day.)

In theory, only the measurement of $\hat{a}_{33} = \hat{\sigma}_z \otimes \hat{\sigma}_z$ should produce a deterministic outcome, $+1$. The other eight measurements, whose projectors are equatorial on the Bloch sphere, should randomly yield the outcomes $+1$ and -1 with equal probability. We see that this is not quite the case as the measurements have a slight bias toward the $+1$ outcome. Other input states such as $|+1_z\rangle \otimes |+1_z\rangle$, $|+1_x\rangle \otimes |+1_y\rangle$, and $|+1_x\rangle \otimes |+1_x\rangle$ produced very similar data, both for the measurements with nominally random outcomes and for that with a nominally deterministic outcome. We did not systematically try with different entangled input states since we noted that significant errors were introduced already in the state preparation stage, and these errors could not be distinguished from the measurement errors.

Table 2. The counts of the two possible outcomes for each entry in the Peres–Mermin square is given by the two numbers between a pair of double vertical lines. For example, out of 8192 measurements of $\hat{\sigma}_x \otimes \hat{\mathbb{I}}$, we obtained 4440 times $+1$ and 3752 times -1.

$+1$	-1	$+1$	-1	$+1$	-1
4440	3752	4279	3913	4553	3639
4353	3839	4376	3816	4636	3556
4565	3627	4488	3704	7798	394

The fact that the nominally random outcomes do not exhibit an exact fifty–fifty split has multiple causes. The first is intrinsic to the quantum nature of the qubit system and is therefore expected to have the stochasticity of a binomial distribution with success probability of $\frac{1}{2}$. Any spread in the statistics beyond one standard deviation of this binomial distribution can be imputed to classical fluctuations in the experimental setup. We indeed found out that the results varied unpredictably from day to day by more than the expected standard deviation $\sqrt{8192 \times \frac{1}{2} \times \frac{1}{2}} \approx 45.3$ from the expected mean $8192 \times \frac{1}{2} = 4096$. As for the bias towards $+1$ outcomes, a systematic error would be its most plausible explanation. It is estimated at about 8% for the nominally random measurements and about 5% for the nominally deterministic ones. Physically, this could be explained by the fact that the states $|+1\rangle$ and $|-1\rangle$ were implemented by the IBM computer as transmon qubits in the ground and excited state, respectively. Through energy-dissipation, the latter will relax to the former, while the former cannot undergo any further dissipation, only dephasing—hence the lopsidedness of the qubits in favour of the ground state $|+1\rangle$.

4.2. The compatibility loophole

The Peres–Mermin setup requires that the two-qubit system undergoes three successive, commuting measurements. These measurements—or at least the first two of them—must be of the quantum non-demolitional type [37]. (A demolitional measurement would otherwise leave no state to subsequently measure.) Moreover, the outcomes of the three individual measurements must not be known, only their overall product should. It is only after the third measurement that this product measurement result shall be read out. To accomplish this, we accumulatively encode the three-measurement product in the state of an ancillary qubit that is only to be read out *a posteriori*. The downside of the approach is that it only becomes possible to verify that the sequentially executed measurements do indeed commute by explicitly permuting their order (i.e., by re-programming the permuted sequences in the quantum computer). This in turn opens up for the so-called compatibility loophole whereby contexts can no longer be assumed to be sets of no-disturbance [38, 39].

To show that a successive readout of the individual measurements in a context will not result in the same measurement statistics as when only the final product result is read out, consider the following example: suppose we prepare two particles in spin up in the y-direction (of the Bloch sphere) and we want to measure them in context $\mathcal{M}_{c_3}$. The first measurement of their spin in the x-direction will then either be $+1$ and collapse the state to either $\uparrow\uparrow$ or $\downarrow\downarrow$ (in the x-direction), or be -1 and collapse the state to either $\uparrow\downarrow$ or $\downarrow\uparrow$ (still in the x-direction). Subsequently measuring the spins in the y-direction would again randomly produce either the result $+1$ or -1, uncorrelated to the first measurement. After the y-measurement, the state would be collapsed along the y-axis. The final measurement in the z-direction would likewise produce either the result $+1$ or -1, yet again uncorrelated to the previous measurement outcomes. Hence, the product of the measurements read out individually in context $\mathcal{M}_{c_3}$ would

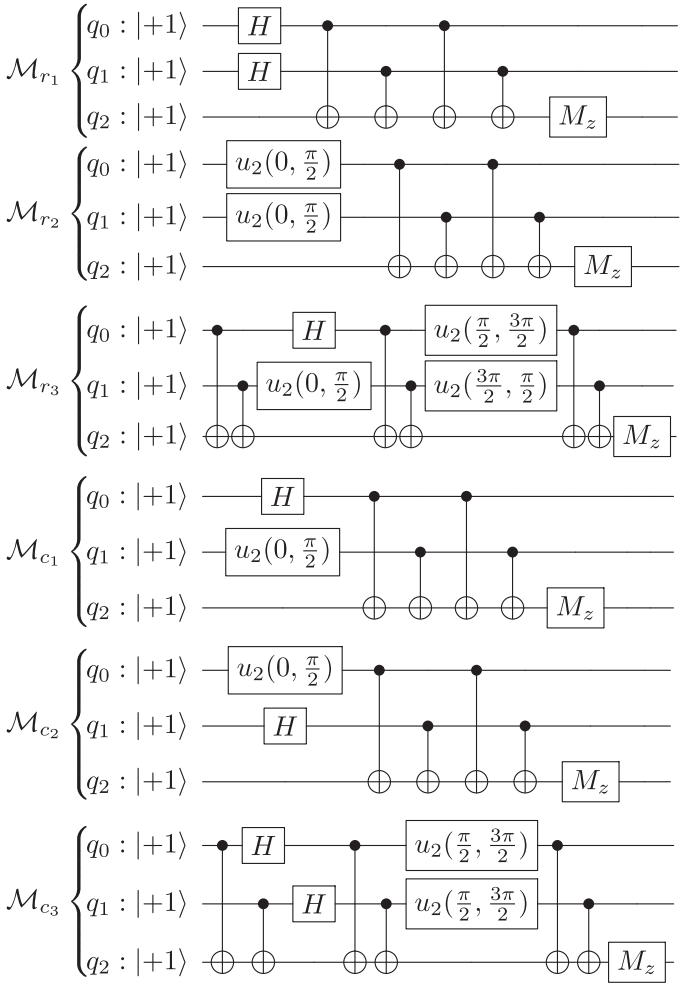

Figure 7. Quantum circuits in the z-basis for the six contexts. The gates used include the unitary rotation $u(\phi, \lambda)$ of the qubit which is composed of a λ-rotation around z, a $\frac{\pi}{2}$-rotation around y, and a ϕ-rotation around x. Also involved is the Hadamard gate H, which flips the poles of the Bloch sphere into two antipodal points on the equator, as well as the controlled NOT gate—symbolized by the vertical links—which effectively serves to entangle the qubits. (An overview of quantum gates is available in [35].)

either be $+1$ or -1 with equal probability, contrary to the prediction of the Peres–Mermin square, and contrary to our experimental results.

4.3. Quantum circuitry

Let us now describe how the experiment is performed—or rather, programmed—on the quantum computer. The six sequence measurements are implemented as the circuits of figure 7. Each circuit is governed by a short computer program that specifies the preparation, processing, and measurement stages [40]. Although the order of the measurements is theoretically irrelevant, the unequal performance of the qubits and the dependence of the gate operations on

Table 3. Breakdown of the counts $n^{(\pm)}$ of the measurement outcomes ± 1 in each of the six contexts.

	$\mathcal{M}_{r_1}$	$\mathcal{M}_{r_2}$	$\mathcal{M}_{r_3}$	$\mathcal{M}_{c_1}$	$\mathcal{M}_{c_2}$	$\mathcal{M}_{c_3}$
$n^{(+)}$	7943	7731	7506	7813	7851	2033
$n^{(-)}$	249	461	686	379	341	6159

their order compelled us to optimize the circuit design. This entailed, among other things, that we minimize the number of gates required since each one comes at a cost of decoherence. For example, we found out that we obtained worse results when the third row of the square was implemented as $\hat{\sigma}_x \otimes \hat{\sigma}_x \cdot \hat{\sigma}_y \otimes \hat{\sigma}_y \cdot \hat{\sigma}_z \otimes \hat{\sigma}_z$ instead of in the reverse order as the latter would otherwise requires more gates[6]. The same could be said of context $\mathcal{M}_{c_3}$, which is also vulnerable to the aforementioned compatibility loophole in that the order of its measurements is *effectively* detrimental to the measurement statistics. The first two rows and first two columns, however, do not suffer from this problem.

4.4. Results and analysis

The best set of results was obtained using the input state $|+1_z\rangle \otimes |-1_z\rangle$. It consists of $n = 8192$ measurements in each of the six contexts, of which $n^{(-)}$ returned the readout -1 while the remaining $n^{(+)}$ returned $+1$ (table 3). The experimental measurement vector can then be shown to be

$$\vec{\mu}^{(\mathrm{IBM})} = \begin{bmatrix} +0.939 \\ +0.887 \\ +0.833 \\ +0.907 \\ +0.917 \\ -0.504 \end{bmatrix} \pm \begin{bmatrix} 3.79 \\ 5.09 \\ 6.12 \\ 4.64 \\ 4.41 \\ 9.55 \end{bmatrix} \cdot 10^{-3}. \tag{31}$$

Each component of $\vec{\mu}^{(\mathrm{IBM})}$, corresponding to some context c, is given by

$$\mu_c^{(\mathrm{IBM})} = \frac{1}{n}\left[(+1) \times n_c^{(+)} + (-1) \times n_c^{(-)}\right] \pm \delta_c, \tag{32}$$

where we have included the standard error of the mean

$$\delta_c = 2\sqrt{\frac{1}{n}\frac{n_c^{(+)}}{n}\frac{n_c^{(-)}}{n}}, \tag{33}$$

as the standard deviation divided by $\sqrt{n}$.

Using the SciPy Python library, we implemented Delaunay triangulation to compute equation (16), and found out that, indeed, our measurement vector lies outside the real hull at an Euclidean distance of

$$\mathcal{D}^{(\mathrm{R})}_{\vec{\mu}^{(\mathrm{IBM})}} \approx 0.4029 \tag{34}$$

[6] We found the former sequence to perform equally well as $\hat{\sigma}_x \otimes \hat{\sigma}_x \cdot \hat{\sigma}_z \otimes \hat{\sigma}_z \cdot \hat{\sigma}_y \otimes \hat{\sigma}_y$. This is predictable since they both have the same number of gates.

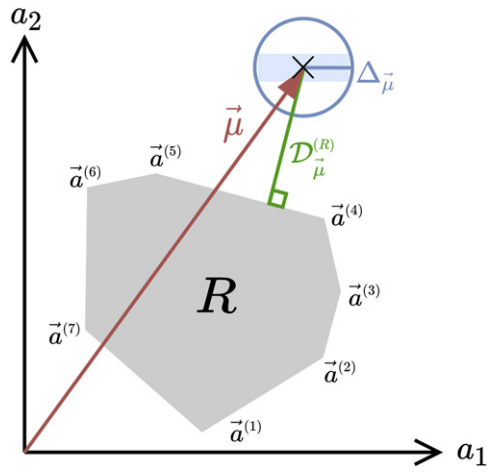

Figure 8. Schematic of a real convex hull R in the a_1a_2-plane supported by the 2D vectors $\vec{a}^{(1)}$ through $\vec{a}^{(7)}$. The point symbolized by $\times$ represents the measurement outcome $\vec{\mu}$ and $\mathcal{D}^{(\mathrm{R})}_{\vec{\mu}}$ is its closest distance to R. The width (height) of the light blue rectangle defines two standard deviations of the measurement in the a_1 (a_2) direction. This rectangle is inscribed in a standard deviation hypersphere (here, the circle in 2D) whose radius $\Delta_{\vec{\mu}}$ is the worst case estimate of the measurement outcome resulting from one standard deviation error in each direction in the plane.

from its surface [41]. To check that this result is statistically significant, we formed a six-dimensional 'hypersphere of uncertainty' around $\vec{\mu}^{(\mathrm{IBM})}$ based on the standard errors of our measurements (figure 8). The radius of this sphere is

$$\Delta_{\vec{\mu}^{(\mathrm{IBM})}} = \sqrt{\vec{\delta}\cdot\vec{\delta}} \approx 0.0145, \tag{35}$$

namely $0.4029/0.0145 \approx 27.8$ times smaller than $\mathcal{D}^{(\mathrm{R})}_{\vec{\mu}^{(\mathrm{IBM})}}$. We can therefore be confident that our measurement results violate the realistic and noncontextuality hypothesis by at least 27 standard errors.

5. Discussion and outlook

This article aims to render the notions of contextuality and violations of realism more accessible to the non-specialist. We presented from the ground up a toy example that illustrates how contextuality is an interpretation for measurement combinations that cannot logically result from the assumption of pre-existing physical attributes. We then drew the parallel between our toy example and an already well-established quantum setup, namely the Peres–Mermin square. The choice of the Peres–Mermin square is owed to its compactness and elegance in that it works deterministically and does not demand any elaborate quantum state engineering. Its only requirement is the processing and measurement of qubits, which makes it ideal for experimentation on the emergent quantum computer platforms such as the IBM Quantum Experience. Such platforms are an easy-to-use playground for experiments which only a decade ago would have required dedicated laboratories. Our results, which confirmed the nonclassicality of the Peres–Mermin scheme by over 27 standard errors, are a testimony to the credibility of online quantum computers as viable tools in fundamental research.

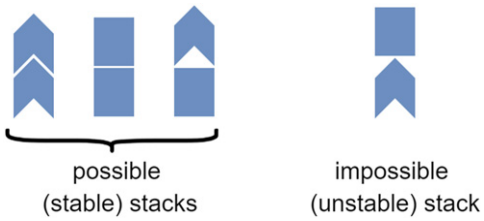

Figure 9. Although there exists $2^2 = 4$ combinations for pairing either chevrons and rectangles, only three of them can be stacked in a stable equilibrium.

This said, we note that our metric for the violation of realism, $\mathcal{D}^{(\mathrm{R})}_{\vec{\mu}^{(\mathrm{IBM})}} \approx 0.4029$, though satisfactory in terms of disproving the null hypothesis beyond the margin of error $\Delta_{\vec{\mu}^{(\mathrm{IBM})}} \approx 0.0145$, is in fact barely half what it could have been in theory, namely $\mathcal{D}^{(\mathrm{R})}_{\vec{\mu}^{(\mathrm{Q})}} = \sqrt{2/3}$. This is because the so-called NISQ computers we have used are indeed still too noisy. In particular, we found that the decoherence of the qubits, the non-ideal gate fidelities, and the high readout errors had to be compensated by careful circuit design and several re-runs of the program. Up to 30% of our program executions were too affected by noise to indicate any non-classicality in their results. This was compounded by unpredictable systematic fluctuations from one day to the next which are presumably due to tune-ups of the hardware. Nonetheless, we expect that all these nuisances will be mitigated as the technology steadily improves [42, 43].

Finally, in order to place our work in a larger context, we should clarify that our discussion assumed the Copenhagen interpretation. For completeness, alternative interpretations, or analogies with classical systems that exhibit 'quantum behaviour', should however be kept in mind [44–47].

Acknowledgments

This work was carried out with the support of the Swedish Research Council under Grant 2014-7869-110765-32 and the Wallenberg Center for Quantum Technology. We acknowledge use of the IBM Quantum Experience. The views expressed are those of the authors and do not reflect the official policy or position of IBM or the IBM Quantum Experience team.

Appendix A. Convex hulls from ontic states (an intuitive example)

This section illustrates in a self-contained and non-mathematical way how ontic (i.e., real) states give rise to the so-called convex hulls. To this end, imagine we have several instances of two possible shapes—rectangles and chevrons—which we would like to group into pairs where one shape is stacked on top of the other. As is illustrated in figure 9, although we have $2^2 = 4$ combinations for these stacks, only three of them are physically possible in the sense that the fourth cannot exist: a rectangle cannot rest in stable equilibrium on top of a chevron.

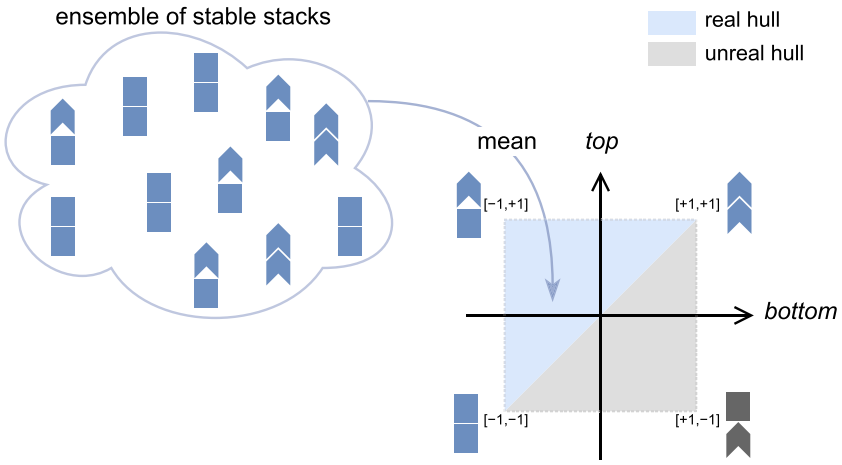

Figure 10. Any conceivable ensemble of pairwise stacks averages to a point on the triangle delimited by $[-1,-1]$, $[-1,+1]$, and $[+1,+1]$. In the particular case illustrate here, we have $\mu = (5[-1,-1] + 2[+1,+1] + 4[-1,+1])/11 \approx [-0.6, 0.1]$.

Without loss of generality, one can represent rectangles and chevrons with -1 and $+1$ respectively such that the four combinations of pairwise stacks can be laid on a two-dimensional plane where the abscissa represents the bottom shape and the ordinate represents the top shape. If we then have an ensemble of such pairwise stacks, we expect that the mean position of elements in the ensemble lies somewhere on the triangle delimited by $[-1,-1]$, $[-1,+1]$, and $[+1,+1]$. This triangle, which charts the realm of possibilities as a two-dimensional surface, is what is referred to as the real convex hull (figure 10). The complementary triangle delimited by $[-1,-1]$, $[+1,+1]$ and the 'impossible' corner $[+1,-1]$ (i.e., rectangle on top of chevron) is called the unreal convex hull in that no point within could possibly result from an ensemble of stable stacks.

Appendix B. Derivation of the third column of the Peres–Mermin square

The definition of tensor products $\hat{A}^{(\mathrm{L})} \otimes \hat{A}^{(\mathrm{R})}$ is that every entry $a_{ij}^{(\mathrm{L})}$ of the left-hand operand is multiplied by the whole right-hand operand. For example, for two-dimensional matrices, we have

$$\begin{aligned}\hat{A}^{(\mathrm{L})} \otimes \hat{A}^{(\mathrm{R})} &= \begin{bmatrix} a_{11}^{(\mathrm{L})} & a_{12}^{(\mathrm{L})} \\ a_{21}^{(\mathrm{L})} & a_{22}^{(\mathrm{L})} \end{bmatrix} \otimes \begin{bmatrix} a_{11}^{(\mathrm{R})} & a_{12}^{(\mathrm{R})} \\ a_{21}^{(\mathrm{R})} & a_{22}^{(\mathrm{R})} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}^{(\mathrm{L})} a_{11}^{(\mathrm{R})} & a_{11}^{(\mathrm{L})} a_{12}^{(\mathrm{R})} & a_{12}^{(\mathrm{L})} a_{11}^{(\mathrm{R})} & a_{12}^{(\mathrm{L})} a_{12}^{(\mathrm{R})} \\ a_{11}^{(\mathrm{L})} a_{21}^{(\mathrm{R})} & a_{11}^{(\mathrm{L})} a_{22}^{(\mathrm{R})} & a_{12}^{(\mathrm{L})} a_{21}^{(\mathrm{R})} & a_{12}^{(\mathrm{L})} a_{22}^{(\mathrm{R})} \\ a_{21}^{(\mathrm{L})} a_{11}^{(\mathrm{R})} & a_{21}^{(\mathrm{L})} a_{12}^{(\mathrm{R})} & a_{22}^{(\mathrm{L})} a_{11}^{(\mathrm{R})} & a_{22}^{(\mathrm{L})} a_{12}^{(\mathrm{R})} \\ a_{21}^{(\mathrm{L})} a_{21}^{(\mathrm{R})} & a_{21}^{(\mathrm{L})} a_{22}^{(\mathrm{R})} & a_{22}^{(\mathrm{L})} a_{21}^{(\mathrm{R})} & a_{22}^{(\mathrm{L})} a_{22}^{(\mathrm{R})} \end{bmatrix}. \end{aligned} \tag{36}$$

Furthermore, when involved in matrix multiplication, tensor products have the property that

$$\left(\hat{A}^{(\mathrm{R})} \otimes \hat{A}^{(\mathrm{L})}\right) \cdot \left(\hat{B}^{(\mathrm{R})} \otimes \hat{B}^{(\mathrm{L})}\right) = \left(\hat{A}^{(\mathrm{R})} \cdot \hat{B}^{(\mathrm{R})}\right) \otimes \left(\hat{A}^{(\mathrm{L})} \cdot \hat{B}^{(\mathrm{L})}\right). \tag{37}$$

Let us now derive the sequence of operations associated with context $\mathcal{M}_{c_3}$, i.e., the third column of the Peres–Mermin square. We have

$$\begin{aligned}
(\hat{\sigma}_x \otimes \hat{\sigma}_x) \cdot (\hat{\sigma}_y \otimes \hat{\sigma}_y) \cdot (\hat{\sigma}_z \otimes \hat{\sigma}_z) &= \{\text{equation (37)}\} \\
&= (\hat{\sigma}_x \cdot \hat{\sigma}_y \cdot \hat{\sigma}_z) \otimes (\hat{\sigma}_x \cdot \hat{\sigma}_y \cdot \hat{\sigma}_z) \\
&= \{\hat{\sigma}_x \cdot \hat{\sigma}_y = i\hat{\sigma}_z\} \\
&= (i\hat{\sigma}_z \cdot \hat{\sigma}_z) \otimes (i\hat{\sigma}_z \cdot \hat{\sigma}_z) \\
&= \left\{\hat{\sigma}_z \cdot \hat{\sigma}_z = \hat{\mathbb{I}}_2\right\} \\
&= i\hat{\mathbb{I}}_2 \otimes i\hat{\mathbb{I}}_2 \\
&= \{\text{equation (36)}\} \\
&= -\hat{\mathbb{I}}_4.
\end{aligned} \tag{38}$$

Hence, the (doubly-degenerate) eigenvalue of measurements in $\mathcal{M}_{c_3}$ is -1.

ORCID iDs

Amine Laghaout https://orcid.org/0000-0001-7891-4505
Gunnar Björk https://orcid.org/0000-0002-2082-9583

References

[1] Einstein A, Podolsky B and Rosen N 1935 Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47** 777–80
[2] Bell J S 1964 On the Einstein–Podolsky–Rosen paradox *Physics* **1** 195–200
[3] Kochen S and Specker E P 1975 *The Problem of Hidden Variables in Quantum Mechanics* (Berlin: Springer) pp 293–328
[4] Huang Y-F, Li C-F, Zhang Y-S, Pan J-W and Guo G-C 2003 Experimental test of the Kochen–Specker theorem with single photons *Phys. Rev. Lett.* **90** 250401
[5] Bartosik H *et al* 2009 Experimental test of quantum contextuality in neutron interferometry *Phys. Rev. Lett.* **103** 040403
[6] Kirchmair G, Zähringer F, Gerritsma R, Kleinmann M, Gühne O, Cabello A, Blatt R and Roos C F 2009 State-independent experimental test of quantum contextuality *Nature* **460** 494–7
[7] Liu B H *et al* 2009 Experimental demonstration of quantum contextuality with nonentangled photons *Phys. Rev.* A **80** 044101
[8] Moussa O, Ryan C A, Cory D G and Laflamme R 2010 Testing contextuality on quantum ensembles with one clean qubit *Phys. Rev. Lett.* **104** 160501
[9] Amselem E, Rådmark M, Bourennane M and Cabello A 2009 State-independent quantum contextuality with single photons *Phys. Rev. Lett.* **103** 160405
[10] Hasegawa Y, Durstberger-Rennhofer K, Sponar S and Rauch H 2011 Kochen–Specker theorem studied with neutron interferometer *Nucl. Instrum. Methods Phys. Res.* A **634** S21–4
[11] Lapkiewicz R, Li P, Schaeff C, Langford N K, Ramelow S, Wieśniak M and Zeilinger A Experimental non-classicality of an indivisible quantum system *Nature* **474** 490–3

[12] Zu C, Wang Y-X, Deng D-L, Chang X-Y, Liu K, Hou P-Y, Yang H-X and Duan L-M 2012 State-independent experimental test of quantum contextuality in an indivisible system *Phys. Rev. Lett.* **109** 150401
[13] Dogra S, Dorai K and Arvind 2016 Experimental demonstration of quantum contextuality on an NMR qutrit *Phys. Lett.* A **380** 1941–6
[14] Cañas G *et al* 2016 Experimental demonstration of the connection between quantum contextuality and graph theory *Phys. Rev.* A **94** 012337
[15] Crespi A *et al* 2017 Single-photon quantum contextuality on a chip *ACS Photon.* **4** 2807–12
[16] Xiao Y *et al* 2018 Experimental observation of quantum state-independent contextuality under no-signaling conditions *Opt. Express* **26** 32–50
[17] Zhang A *et al* 2019 Experimental test of contextuality in quantum and classical systems *Phys. Rev. Lett.* **122** 080401
[18] Kleinmann M, Gühne O, Portillo J R, Larsson J-Å and Cabello A 2011 Memory cost of quantum contextuality *New J. Phys.* **13** 113011
[19] Hensen B *et al* 2015 Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres *Nature* **526** 682–6
[20] Peres A 1990 Incompatible results of quantum measurements *Phys. Lett.* A **151** 107–8
[21] Mermin N D 1990 Simple unified form for the major no-hidden-variables theorems *Phys. Rev. Lett.* **65** 3373–6
[22] Mermin N D 1993 Hidden variables and the two theorems of John Bell *Rev. Mod. Phys.* **65** 803–15
[23] Alsina D and Latorre J I 2016 Experimental test of Mermin inequalities on a five-qubit quantum computer *Phys. Rev.* A **94** 012314
[24] Devitt S J 2016 Performing quantum computing experiments in the cloud *Phys. Rev.* A **94** 032329
[25] García-Martín D and Sierra G 2018 Five experimental tests on the five-qubit IBM quantum computer *J. Appl. Math. Phys.* **06** 1460
[26] Smart S E, Schuster D I and Mazziotti D A 2019 Experimental data from a quantum computer verifies the generalized Pauli exclusion principle *Commun. Phys.* **2** 11
[27] Spekkens R W 2005 Contextuality for preparations, transformations, and unsharp measurements *Phys. Rev.* A **71** 052108
[28] Schmid D, Selby J H and Spekkens R W 2021 Unscrambling the omelette of causation and inference: the framework of causal-inferential theories (arXiv:2009.03297)
[29] Cavalcanti E G 2018 Classical causal models for Bell and Kochen–Specker inequality violations require fine-tuning *Phys. Rev.* X **8** 021018
[30] Abramsky S 2020 Contextuality: at the borders of paradox (arXiv:2011.04899)
[31] Cabello A, Severini S and Winter A 2014 Graph-theoretic approach to quantum correlations *Phys. Rev. Lett.* **112** 040401
[32] Scarani V, Gisin N, Brunner N, Masanes L, Pino S and Acín A 2006 Secrecy extraction from no-signaling correlations *Phys. Rev.* A **74** 042339
[33] Cabello A 2008 Experimentally testable state-independent quantum contextuality *Phys. Rev. Lett.* **101** 210401
[34] La Cour B R 2009 Quantum contextuality in the Mermin–Peres square: a hidden-variable perspective *Phys. Rev.* A **79** 012102
[35] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)
[36] IBM quantum experience https://quantum-computing.ibm.com/
[37] Braginsky V B, Vorontsov Y I and Thorne K S 1980 Quantum nondemolition measurements *Science* **209** 547–57
[38] Gühne O *et al* 2010 Compatibility and noncontextuality for sequential measurements *Phys. Rev.* A **81** 022121
[39] Kujala J V, Dzhafarov E N and Larsson J-Å 2015 Necessary and sufficient conditions for an extended noncontextuality in a broad class of quantum mechanical systems *Phys. Rev. Lett.* **115** 150401
[40] Python scripts for programming the Peres–Mermin square on the IBM quantum experience https://github.com/AltayDikme-R/Mermin-Peres-Magic-Square (accessed 2 August 2021)
[41] Laghaout A 2021 Object-oriented framework for simulating quantum contextuality https://github.com/laghaout/contextuality (accessed 10 August 2021)
[42] Arute F *et al* 2019 Quantum supremacy using a programmable superconducting processor *Nature* **574** 505–10

[43] Place A P M *et al* 2021 New material platform for superconducting transmon qubits with coherence times exceeding 0.3 milliseconds *Nat. Commun.* **12** 1779
[44] Wallentin F 2021 Contextuality in classical physics and its impact on the foundations of quantum mechanics *Entropy* **23** 968
[45] Couder Y and Fort E 2012 Probabilities and trajectories in a classical wave-particle duality *J. Phys.: Conf. Ser.* **361** 012001
[46] Dragoman D and Dragoman M 2004 *Quantum-Classical Analogies* (Berlin: Springer)
[47] Abad L V 2012 Principles of classical statistical mechanics: a perspective from the notion of complementarity *Ann. Phys., NY* **327** 1682–93