Quantum Advantage from Sequential-Transformation Contextuality

Shane Mansfield, Elham Kashefi

Quantum Advantage from Sequential-Transformation Contextuality

Shane Mansfield[1, *] and Elham Kashefi[1, 2]

[1] *Sorbonne Université, CNRS, Laboratoire d'Informatique de Paris 6, F-75005 Paris, France*

[2] *School of Informatics, University of Edinburgh, 10 Crichton St, Edinburgh EH8 9AB, United Kingdom*

(Dated: December 11, 2018)

We introduce a notion of contextuality for transformations in sequential contexts, distinct from the Bell-Kochen-Specker and Spekkens notions of contextuality. Within a transformation-based model for quantum computation we show that strong sequential-transformation contextuality is necessary and sufficient for deterministic computation of non-linear functions if classical components are restricted to mod2-linearity and matching constraints apply to any underlying ontology. For probabilistic computation, sequential-transformation contextuality is necessary and sufficient for advantage in this task and the degree of advantage quantifiably relates to the degree of contextuality.

Contextuality is a key non-classical phenomenon exhibited by quantum systems, which was first considered by Bell [1] and by Kochen and Specker [2]. It has been the subject of renewed interest recently as a range of results have established it to be the essential ingredient for enabling quantum advantages over classical implementations of a variety of informatic tasks [3–5], simulation of quantum processes [6], and for enabling universal quantum computing [7–11][70]. A broader notion of contextuality due to Spekkens [12] has also been shown to be essential to quantum advantages relating to state discrimination and one-way communication protocols [13–17]. However, questions remain over which forms of contextuality provide advantage in which precise settings [18] and whether existing notions of contextuality are sufficient to account for all instances of quantum advantage. For example, there exist a variety of advantages achievable with a single qubit [19–21], where Bell-Kochen-Specker (BKS) contextuality cannot arise [2, 22] and to which there is no apparent link to the Spekkens version. This raises the important question of which non-classical feature could be at play if not contextuality of these kinds.

We introduce a notion of contextuality for transformations performed in sequential contexts that is inequivalent to the notion of transformation contextuality introduced by Spekkens [23]. It is necessarily present in a recently discovered form of quantum advantage in shallow circuits [24]. We will show, via a Mermin-style [25, 26] parity argument, that it is also crucial in enabling increased computational power in the single qubit example of [21]. The setting for that example is a transformation-based model of quantum computing, which we call here $l2$-TBQC, that was shown to be useful in achieving secure delegated computing. In the model, a classical control computer, whose power is limited to mod2-linear computation, may interact with a quantum resource, by which its computational power may be enhanced. As with the analogous measurement-based model, $l2$-MBQC [4], which was the setting for the results of [3–5], it can provide a useful tool for probing the roots of quantum advantage. In this setting, we show more generally that sequential-transformation contextuality is necessary and sufficient to enable advantage in the task of probabilistically computing any non-linear function whenever classical ontologies are required to respect the computational assumptions. Moreover, the degree of contextuality can be related to the probability of success, and in particular, strong (i.e., maximal) contextuality is necessary for deterministic computation of any non-linear function.

Our results trace an arc that parallels developments relating BKS contextuality to quantum advantage in $l2$-MBQC: Anders and Browne provided an example in which a contextual resource is sufficient for the computation of a particular non-linear function [3]; Raussendorf then proved that strong contextuality is necessary for any deterministic non-linear computation [4], as initially observed by Hoban *et al.* for non-adaptive $l2$-MBQC [27] based on an early version of [4]; he also showed that contextuality is necessary for quantum advantage in the task of probabilistically computing any non-linear function; this latter result was later sharpened to show more precisely how the degree of contextuality as measured by the contextual fraction relates to probability of success in [5]. Our results set the stage for further investigation of how sequential-transformation contextuality may relate to quantum advantages, speedups and the onset of universality in other settings, as the results of [7–10] do for BKS contextuality.

Ontological models.— Quantum theory exhibits a number of apparently non-intuitive features. Crucially, in many cases there exist no-go theorems that establish that there is no way these features can be explained away by recourse to any deeper or more complete theory that would obey certain classical intuitions [28]. Some such non-classical features are non-locality [29] (BKS) contextuality [1, 2] [71], forms of preparation and transformation contextuality [23], while others relate to macrorealism [30–33] and the ontic nature of the quantum state [34–41]. A convenient formalism for treating such theorems is that of ontological models, which we briefly set out next. Note that in this work when we speak of ontological models we will not be assuming any additional features beyond what is explicitly set out below (e.g., of the kind present in [23]).

The central component is an ontic state space Λ, comprising the states of a hypothetical underlying theory. Preparation of a quantum state ρ results in an ontic state sampled according to a probability distribution d_ρ on Λ [72]. In the simplest case, a quantum transformation U corresponds to a measurable function $f_U : \Lambda \to \Lambda$. For consistency we require that $f_{U*}d_\rho = d_{U\rho U^\dagger}$, where the left-hand side is the push forward of d_ρ along f_U, defined by $f_{U*}d_\rho(\lambda) = d_\rho[f_U^{-1}(\lambda)]$. We also require that the function corresponding to the identity operator simply maps each ontic state to the δ function centered on that state, ensuring that $f_{\mathbb{1}*}d_\rho = d_\rho$ for all preparations ρ. In particular, the requirements entail that unitaries correspond to invertible functions. A quantum measurement M corresponds to a function $\xi_M : \Lambda \to P(O)$ which assigns to each ontic state a probability distribution over the set of outcomes O. For any combination of preparation, transformation, and measurement, the ontological theory predicts that the empirical statistics, $e_{\rho,U,M} \in P(O)$, are given by

$$e_{\rho,U,M} = \sum_{\lambda\in\Lambda} d_\rho(\lambda)\, \xi_M(f_U(\lambda)) \,. \tag{1}$$

In fact, our results apply more generally to ontological models in which transformations may correspond to stochastic mixtures of measurable functions. However, we will see shortly that, for our present purposes, since such an ontological model can always be expressed as a convex decomposition of ones in which transformations are deterministic, it will suffice to establish no-go properties for those with deterministic transformations. No-go theorems arise when it is found that ontological models satisfying some additional, perhaps "classical", assumptions are unable to realise the empirical predictions of quantum theory.

(Non-)contextuality.— In the BKS sense, non-contextuality is an assumption of classicality that applies when certain finite sets of compatible measurements may be performed jointly in contexts. It requires that for each valid context C compatibility is reflected at the ontological level through factorisability of the joint measurement function $\xi_C : \Lambda \to P(O^{|C|})$; i.e.,

$$\xi_C = \prod_{M\in C} \xi_M \,. \tag{2}$$

Implicit in this is the crucial requirement that, for any measurement M occurring in contexts C and C', its ontological representation ξ_M is context independent; i.e.,

$$\xi_{M^{(C)}} = \xi_{M^{(C')}} \,.$$

This description of non-contextuality via factorisability is equivalent to the description in terms of global valuations that may be more familiar to some readers [42].

Next, we mention some specific instances arising from Spekkens' general notion of non-contextuality [23]. Measurement non-contextuality in the explicit sense treated in the no-go results of [23] relaxes (2) to the weaker requirement that

$$\xi_C|_M = \xi_M \,,$$

for all M and C such that $M \in C$, where $\xi_C|_M$ denotes the marginalisation of ξ_C to M.

Transformation and preparation non-contextuality in the explicit sense treated in the no-go results of [23] takes as context any convex decomposition of a given transformation or preparation. This has an operational motivation. Suppose, as a concrete example, that some transformation T admits the following unitary decompositions:

$$T = \frac{1}{2}U_a + \frac{1}{2}U_A \,, \tag{C}$$

$$T = \frac{1}{3}U_a + \frac{1}{3}U_b + \frac{1}{3}U_c \,. \tag{C'}$$

Operationally, context C is "apply U_a or U_A uniformly at random", and context C' is "apply U_a, U_b or U_c uniformly at random"; quantum mechanically the contexts are equivalent. Non-contextuality requires that convex decompositions are reflected at the ontological level; i.e., in this instance,

$$f_T = \frac{1}{2}f_{U_a} + \frac{1}{2}f_{U_A} = \frac{1}{3}f_{U_a} + \frac{1}{3}f_{U_b} + \frac{1}{3}f_{U_c} \,.$$

Again, it is implicit that ontological representations of transformations and preparations are independent of operational context; e.g.,

$$f_{U_a^{(C)}} = f_{U_a^{(C')}} \,.$$

Sequential transformations.— With the preceding versions for comparison, we now introduce a version of non-contextuality for transformations in sequential contexts. It requires that for each finite sequence of transformations, $C = \{U_i\}_{i=1}^t$, sequential composition is reflected at the ontological level; i.e.,

$$f_{U_t\cdots U_1} = f_{U_t} \circ \cdots \circ f_{U_1} \,.$$

It is assumed that the ontological representations of transformations are independent of sequential context; i.e., whenever a transformation U occurs in contexts C and C', it holds that

$$f_{U^{(C)}} = f_{U^{(C')}} \,.$$

When a set of empirical data or predictions cannot be reproduced by an ontological model satisfying this property, it is said to be contextual.

Contextuality in our sense implies that the system of study cannot have an ontology in which transformations correspond to modular, composable operations on ontic states, such that they are well defined independently of which transformations may have been performed previously or will be performed subsequently. Either we must reject the ontological picture entirely or give up on these highly intuitive, classical properties. Note that one plausible, if conspiratorial, mechanism for introducing some contextuality might be through causal dependence on transformations having appeared earlier in the sequence, but even this kind of mechanism is precluded when the transformations being modelled commute.

The constant-depth quantum circuits of [24] provide a concrete example of sequential-transformation contextuality as they can at best be simulated by classical circuits whose depth grows logarithmically in the size of the input. If a modular, non-contextual ontological description of gate transformations at each step in the circuit were possible, then it would give rise to classical circuits for the same task, which would also have constant depth. Connections to quantum advantage in this setting will be investigated in future work; here we focus on examples in a more restricted setting.

Quantification.— An empirical model $e = \{e_C\}$, associates with each context C a distribution over observed outcomes [42]. Similar to [5], given any empirical model and appropriate version of contextuality, we may consider convex decompositions of the form

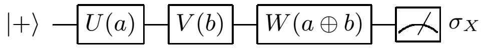
$$e = \omega e^{\mathrm{NC}} + (1-\omega)e' \,, \tag{3}$$

where e^{NC} and e' are also empirical models, and e^{NC} is non-contextual. The maximum value of ω over all such decompositions is the non-contextual fraction of e, written $\mathrm{NCF}(e)$, and correspondingly, the contextual fraction of e is $\mathrm{CF}(e) := 1 - \mathrm{NCF}(e)$ [73]. For BKS contextuality, the contextual fraction corresponds to the maximum achievable normalised violation by e of any generalised Bell inequality [5]. Here, however, we use it to quantify sequential-transformation contextuality. Using the terminology of the hierarchy of BKS contextuality introduced in [42], an empirical model is said to be strongly contextual when $\mathrm{CF}(e) = 1$.

For a given experimental scenario, the set of all the possible e^{NC} is convex, and any extremal point corresponds simply to fixing a deterministic function $f_U : \Lambda \to \Lambda$ for each transformation U featuring in the scenario. Strong contextuality thus arises in the extreme case that no global assignment of deterministic functions to transformations is consistent with even a fraction of the empirical behaviour.

l2-TBQC.— We consider a classical control computer restricted to mod2-linear computation that can interact with a resource, which may be quantum, as follows. The resource is prepared in a fixed state, the control computer may interact with it by means of controlled transformations, then a fixed measurement is performed on the resource and its outcome returned to the control computer. This captures, for example, the single qubit protocols of [21], which were considered for their security features in a setting in which a client delegates certain operations making up an $l2$-TBQC, such as state preparation and measurement, to a server.

Figure 1: The basic single qubit AND protocol from [21].



Note that, independent of the $l2$ restriction, any measurement-based quantum computation [43] can equivalently be expressed as a TBQC, since choice of a measurement setting is equivalent to choice of a transformation prior to a fixed measurement.

An example of an $l2$-TBQC that performs a basic non-linear function, the AND gate on classical input bits a and b,

$$g(a,b) = (a \oplus 1) \otimes (b \oplus 1) \oplus 1 \,,$$

is the following (Fig. 1) [74]. The control computer receives inputs a and b. For the resource, the fixed state is the qubit state $|+\rangle$, the fixed measurement is given by the Pauli operator σ_X, and the controlled transformations are $U(a)$, then $V(b)$, then $W(a \oplus b)$, where

$$\begin{aligned} U(0) &= V(0) = W(0) = I \,, \\ U(1) &= V(1) = W(1) = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/2} \end{pmatrix} . \end{aligned}$$

Notice that all transformations commute. The output of the computation is the measurement outcome interpreted in $\mathbb{Z}_2$, with eigenvalues $+1$ and -1 mapped to 0 and 1, respectively. In terms of complexity classes, access to a qubit quantum resource promotes the computational power from the class $\oplus L$ [44, 45] to P, as with the example in [3] in the setting of $l2$-MBQC.

$\oplus L$-ontology.— Of course, classical computers can perfectly well compute non-linear functions and they also constitute valid non-contextual ontologies. To pose a meaningful computational question about whether a resource may be used to boost power from $\oplus L$ to P, therefore, we will restrict attention to $\oplus L$-ontologies, which we define as follows. Recalling that $\oplus L$ circuits are built entirely of NOT and controlled-NOT (CNOT) gates [45], we will suppose that available transformations are built from these and act on an ontic state space $\mathbb{Z}_2^s$, for some $s \in \mathbb{N}$. In what follows, we will be interested in protocols in which transformations commute. These can already permit efficient solutions to problems for which it is believed there can be no efficient classical solution [46]. For transformations in commutative $\oplus L$ ontologies it holds that, for any transformation U,

$$f_U(\boldsymbol{\lambda}) = (I \oplus A_U)\,\boldsymbol{\lambda} \oplus \boldsymbol{u} \,, \tag{4}$$

where A_U is an $s \times s$ matrix over $\mathbb{Z}_2$ containing only off-diagonal entries and $\boldsymbol{u} \in \mathbb{Z}_2^s$ [see Appendix]. For composition of transformations $\{U_i\}_{i=1}^t$ with ontological representations determined by $\{A_i, \boldsymbol{u}_i\}$ it holds that

$$f_{U_t} \circ \cdots \circ f_{U_1}(\boldsymbol{\lambda}) = \boldsymbol{\lambda} \oplus \bigoplus_{i=1}^{t} A_i \boldsymbol{\lambda} \oplus \bigoplus_{i=1}^{t} \boldsymbol{u}_i \,.$$

A dichotomic measurement in an $\oplus L$ ontology can most generally be described by a transformation followed by output of the bit value of a fixed entry j of the final ontic state vector; i.e. $\boldsymbol{\lambda}' \cdot \boldsymbol{\delta}$ where $\boldsymbol{\lambda}', \boldsymbol{\delta} \in \mathbb{Z}_2^s$ are the post-transformation ontic state and the vector with jth entry 1 and 0's elsewhere, respectively.

RESULTS

Proposition 1. *Any commutative $\oplus L$-ontological realisation of the* AND *l2-TBQC is transformation contextual.*

Proof. Suppose that preparation results in an initial ontic state $\boldsymbol{\lambda} \in (\mathbb{Z}_2)^s$. From (1), non-contextual realisation of the protocol requires Eqs (5–8) to be satisfied. These describe evaluation of the computation for the four possible sequential contexts, where ontological representations of $U(k)$, $V(k)$, and $W(k)$, with $k \in \{0,1\}$, are determined through Eq. (4) by $\{A_U(k), \boldsymbol{u}(k)\}$, $\{A_V(k), \boldsymbol{v}(k)\}$, and $\{A_W(k), \boldsymbol{w}(k)\}$, respectively, and of the transformation component of the measurement by $\{A_M, \boldsymbol{m}\}$,

$$[\boldsymbol{\lambda} \oplus A_U(0)\boldsymbol{\lambda} \oplus A_V(0)\boldsymbol{\lambda} \oplus A_W(0)\boldsymbol{\lambda} \oplus A_M\boldsymbol{\lambda} \oplus \boldsymbol{u}(0) \oplus \boldsymbol{v}(0) \oplus \boldsymbol{w}(0) \oplus \boldsymbol{m}] \cdot \boldsymbol{\delta} = 0 \,, \tag{5}$$

$$[\boldsymbol{\lambda} \oplus A_U(0)\boldsymbol{\lambda} \oplus A_V(1)\boldsymbol{\lambda} \oplus A_W(1)\boldsymbol{\lambda} \oplus A_M\boldsymbol{\lambda} \oplus \boldsymbol{u}(0) \oplus \boldsymbol{v}(1) \oplus \boldsymbol{w}(1) \oplus \boldsymbol{m}] \cdot \boldsymbol{\delta} = 0 \,, \tag{6}$$

$$[\boldsymbol{\lambda} \oplus A_U(1)\boldsymbol{\lambda} \oplus A_V(0)\boldsymbol{\lambda} \oplus A_W(1)\boldsymbol{\lambda} \oplus A_M\boldsymbol{\lambda} \oplus \boldsymbol{u}(1) \oplus \boldsymbol{v}(0) \oplus \boldsymbol{w}(1) \oplus \boldsymbol{m}] \cdot \boldsymbol{\delta} = 0 \,, \tag{7}$$

$$[\boldsymbol{\lambda} \oplus A_U(1)\boldsymbol{\lambda} \oplus A_V(1)\boldsymbol{\lambda} \oplus A_W(0)\boldsymbol{\lambda} \oplus A_M\boldsymbol{\lambda} \oplus \boldsymbol{u}(1) \oplus \boldsymbol{v}(1) \oplus \boldsymbol{w}(0) \oplus \boldsymbol{m}] \cdot \boldsymbol{\delta} = 1 \,. \tag{8}$$

Under the assumption of non-contextuality, the equations are not jointly satisfiable. This can be deduced from the fact that the sum modulo 2 of the right-hand sides is one, whereas the sum of the left-hand sides is zero, since each vector appears an even number of times leading to cancellations. Note that a contextual realisation would permit ontological representations to vary according to context; e.g., $\boldsymbol{u}(0)^{(4)} \neq \boldsymbol{u}(0)^{(5)}$. Contextually, we can always satisfy the equations. The conclusion is that, while $\oplus L$-ontological descriptions are possible, they are necessarily transformation contextual. □

The above proof is similar to Mermin's parity version [25, 26] of the Greenberger-Horne-Shimony-Zeilinger inquality-free argument for non-locality [47, 48], and is an instance of an all-versus-nothing proof of strong contextuality [49], albeit for transformation rather than BKS contextuality.

Proposition 2. *Strong transformation contextuality is necessary for $\oplus L$-ontological realisation of any non-linear commutative l2-TBQC.*

Proof. Let $\boldsymbol{i} \in (\mathbb{Z}_2)^r$ be the input of the computation and $\boldsymbol{\lambda} \in (\mathbb{Z}_2)^s$ be the initial ontic state resulting from the fixed preparation. The control bits $\boldsymbol{k} = \{k_i\}_{i=1}^t$ for the transformations to be performed are linearly determined from the inputs:

$$\boldsymbol{k} = B\boldsymbol{i} \oplus \boldsymbol{c} \,. \tag{9}$$

for some $n \times r$ matrix B and vector $\boldsymbol{c}$ of length r over $\mathbb{Z}_2$. The transformations to be performed in sequence are $\{U_i(k_i)\}_{i=1}^t$. In a non-contextual model, these will have ontological representations determined through Eq. (11) by $\{A_i(k_i), \boldsymbol{u}_i(k_i)\}$. These are necessarily linear in k_i for each i, since entries in A_i and $\boldsymbol{u}_i$ take values in $\mathbb{Z}_2$ and are functionally determined from k_i, but all functions of type $\mathbb{Z}_2 \to \mathbb{Z}_2$ are linear. Deterministic realisation of a function $g : (\mathbb{Z}_2)^r \to \mathbb{Z}_2$ by a non-contextual ontological model requires that, for all inputs $\boldsymbol{i}$,

$$g(\boldsymbol{i}) = \left[\left(I \oplus \bigoplus_{i=1}^{t} A_{U_i}(k_i) \oplus A_M\right)\boldsymbol{\lambda} \oplus \bigoplus_{i=1}^{n} \boldsymbol{u}_i(k_i) \oplus \boldsymbol{u}_M\right] \cdot \boldsymbol{\delta} \,.$$

This is a linear function since right-hand side expression is linear in $\boldsymbol{k}$, and in turn $\boldsymbol{k}$ is linear in $\boldsymbol{i}$, by Eq. (9).

In a contextual model we could allow the ontological representations to have context dependence; i.e., $\{A_i(\boldsymbol{k}), \boldsymbol{u}_i(\boldsymbol{k})\}$. In this case the entries of the respective matrices and vectors are determined by functions $\mathbb{Z}_2^n \to \mathbb{Z}_2$, which can introduce non-linearity.

Moreover, if any fraction p of the empirical behaviour can be described non-contextually, then with an average probability over all possible inputs of at least p, the $l2$-TBQC computes some linear function. Therefore deterministic computation of a non-linear function requires strong contextuality. □

Given two functions $g, h : \mathbb{Z}_2^r \to \mathbb{Z}_2$, we can define an average distance between these functions as

$$d(g,h) := 2^{-r} \left|\{\boldsymbol{i} \mid g(\boldsymbol{i}) \neq h(\boldsymbol{i})\}\right| .$$

This can be used to measure the degree of non-linearity of any function $g : \mathbb{Z}_2^r \to \mathbb{Z}_2$ as the distance to the closest linear function of that type,

$$\nu(g) := \min\left\{d(g,h) \mid h : \mathbb{Z}_2^r \to \mathbb{Z}_2 \text{ linear}\right\} .$$

Theorem 1. *If a commutative $l2$-TBQC, with resource empirical model e, probabilistically computes a function $g : \mathbb{Z}_2^r \to \mathbb{Z}_2$ with an average failure probability ε over all 2^r possible inputs, then*

$$\varepsilon \geq \mathsf{NCF}(e)\,\nu(g) .$$

Proof. The average probability of success is $p_S := 1 - \varepsilon$. From (3), we can decompose the resource empirical model as

$$e = \mathsf{NCF}(e)\, e^{\mathsf{NC}} + \mathsf{CF}(e)\, e' ,$$

where e' is necessarily strongly contextual. This allows us to similarly decompose the behaviour of the $l2$-TBQC, so that

$$p_S = \mathsf{NCF}(e)\, p_{S,e^{\mathsf{NC}}} + \mathsf{CF}(e)\, p_{S,e'} ,$$

where $p_{S,e^{\mathsf{NC}}}$ and $p_{S,e'}$ are the average probabilities of success that would be associated with resource empirical models e^{NC} and e', respectively. At best, e' enables deterministic computation of g. This leads to the inequalities

$$\begin{aligned} p_S &\leq \mathsf{NCF}(e)\, p_{S,e^{\mathsf{NC}}} + \mathsf{CF}(e) , \\ \varepsilon &\geq \mathsf{NCF}(e)\, \varepsilon_{e^{\mathsf{NC}}} , \end{aligned} \tag{10}$$

where $\varepsilon_{e^{\mathsf{NC}}} = 1 - p_{S,e^{\mathsf{NC}}}$ is the average probability of failure associated with e^{NC}. From the proof of Propositon 2, we know that for a non-contextual resource empirical model any $l2$-TBQC can only compute convex mixtures of linear functions. Thus $\varepsilon_{e^{\mathsf{NC}}} \leq \tilde{\nu}(g)$, which combined with (10) yields the desired inequality. □

Theorem 1 extends Proposition 2, since, in particular it implies that deterministic computation ($\varepsilon = 0$) of a non-linear function [$\nu(g) > 0$] requires strong contextuality [$\mathsf{NCF}(e) = 0$]. The proof here is similar to that of Theorem 3 in [5].

DISCUSSION

The present results highlight the potential of sequential contextuality as a source of quantum advantage of a single qubit over arbitrarily many classical bits for a particular kind of computational task. While the $\oplus L$-ontological assumptions are natural in the particular setting of restricted classical computation that we consider, a direction for future research will be to consider examples of sequential transformation contextuality in less restricted settings, like that of [24], as well as to explore other potential connections to quantum advantage, especially in single qubit systems [19, 20]. It also remains to be seen how the present notion of contextuality can be treated in resource-theoretic frameworks of the kind developed in [5, 50–53]. A related analysis, in terms of irreversibility, of transformation-based protocols is contained in [54], and, in the future, it may be interesting to consider advantages as arising from a combination of these phenomena. From a foundational perspective, in light of the present analysis, the experimental results of [55, 56] could already be said to provide indirect experimental evidence for a kind of sequential transformation contextuality, but this leaves open the possibility for experiments designed specifically to test for the feature, which might also aim to minimise potential issues, such as the detection loophole.

Acknowledgements.— The authors thank Samson Abramsky, Rui Soares Barbosa, Dan Browne, Ulysse Chabaud, Tom Douce, Pierre-Emmanuel Emeriau, Ernesto Galvão, Frédéric Grosshans, Luciana Henaut, Matty Hoban, Aleks Kissinger, Jan-Åke Larsson, Damian Markham, and Anna Pappa for valuable discussions and comments. Early ideas for this work were conceived while S.M. was visiting the Simons Institute for the Theory of Computing at the University of California, Berkeley, as a participant of the Logical Structures in Computation program. This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie Grant Agreement No. 750523.

* Electronic address: shane.mansfield@lip6.fr

[1] J. S. Bell, Reviews of Modern Physics **38**, 447 (1966).

[2] S. Kochen and E. P. Specker, in *The Logico-Algebraic Approach to Quantum Mechanics* (Springer, 1975) pp. 263–276.

[3] J. Anders and D. E. Browne, Physical Review Letters **102**, 050502 (2009).

[4] R. Raussendorf, Physical Review A **88**, 022322 (2013).

[5] S. Abramsky, R. S. Barbosa, and S. Mansfield, Phys. Rev. Lett. **119**, 050504 (2017).

[6] A. Karanjai, J. J. Wallman, and S. D. Bartlett, arXiv preprint arXiv:1802.07744 (2018).

[7] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Nature **510**, 351 (2014).

[8] N. Delfosse, P. A. Guerin, J. Bian, and R. Raussendorf, Physical Review X **5**, 021003 (2015).

[9] H. Pashayan, J. J. Wallman, and S. D. Bartlett, Physical Review Letters **115**, 070501 (2015).

[10] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf, Phys. Rev. Lett. **119**, 120505 (2017).

[11] L. Catani and D. E. Browne, arXiv preprint arXiv:1711.08676 (2017).

[12] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Physical review letters **102**, 010401 (2009).

[13] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, New Journal of Physics **18**, 045003 (2016).

[14] D. Saha, P. Horodecki, and M. Pawłowski, arXiv preprint arXiv:1708.04751 (2017).

[15] D. Schmid and R. W. Spekkens, Physical Review X **8**, 011015 (2018).

[16] D. Saha and A. Chaturvedi, arXiv preprint arXiv:1802.07215 (2018).

[17] S. Ghorai and A. Pan, arXiv preprint arXiv:1806.01194 (2018).

[18] P. Lillystone, J. J. Wallman, and J. Emerson, arXiv preprint arXiv:1802.06121 (2018).

[19] E. Knill and R. Laflamme, Physical Review Letters **81**, 5672 (1998).

[20] E. F. Galvão and L. Hardy, Physical Review Letters **90**, 087902 (2003).

[21] V. Dunjko, T. Kapourniotis, and E. Kashefi, Quantum Information and Computation **16**, 0061 (2016).

[22] A. M. Gleason, Journal of mathematics and mechanics **6**, 885 (1957).

[23] R. W. Spekkens, Physical Review A **71**, 052108 (2005).

[24] S. Bravyi, D. Gosset, and R. Koenig, Science **362**, 308 (2018).

[25] N. D. Mermin, American Journal of Physics **58**, 731 (1990).

[26] N. D. Mermin, Physical Review Letters **65**, 3373 (1990).

[27] M. J. Hoban, E. T. Campbell, K. Loukopoulos, and D. E. Browne, New Journal of Physics **13**, 023014 (2011).

[28] A. Einstein, B. Podolsky, and N. Rosen, Physical review **47**, 777 (1935).

[29] J. S. Bell, Physics **1**, 195 (1964).

[30] A. J. Leggett and A. Garg, Physical Review Letters **54**, 857 (1985).

[31] C. Brukner, S. Taylor, S. Cheung, and V. Vedral, arXiv preprint quant-ph/0402127 (2004).

[32] C. Timpson and O. Maroney, The British Journal for the Philosophy of Science (2013).

[33] J.-M. A. Allen, O. J. E. Maroney, and S. Gogioso, Quantum **1**, 13 (2017).

[34] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Physics Reports **506**, 1 (2011).

[35] M. F. Pusey, J. Barrett, and T. Rudolph, Nature Physics **8**, 475 (2012).

[36] R. Colbeck and R. Renner, Physical Review Letters **108**, 150402 (2012).

[37] R. Colbeck and R. Renner, New Journal of Physics **19**, 013016 (2017).

[38] L. Hardy, International Journal of Modern Physics B **27** (2013).

[39] A. Montina, Modern Physics Letters A **30** (2015).

[40] S. Mansfield, Physical Review A **94**, 042124 (2016).

[41] J.-M. A. Allen, Quantum Studies: Mathematics and Foundations **3**, 161 (2016).

[42] S. Abramsky and A. Brandenburger, New Journal of Physics **13**, 113036 (2011).

[43] R. Raussendorf and H. J. Briegel, Physical Review Letters **86**, 5188 (2001).

[44] C. Damm, Information Processing Letters **36**, 247 (1990).

[45] S. Aaronson and D. Gottesman, Physical Review A **70**, 052328 (2004).

[46] M. J. Bremner, R. Jozsa, and D. J. Shepherd, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* (The Royal Society, 2010) p. rspa20100301.

[47] D. M. Greenberger, M. A. Horne, and A. Zeilinger, in *Bell's theorem, quantum theory and conceptions of the universe* (Springer, 1989) pp. 69–72.

[48] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, American Journal of Physics **58**, 1131 (1990).

[49] S. Abramsky, R. S. Barbosa, K. Kishida, R. Lal, and S. Mansfield, in *24th EACSL Annual Conference on Computer Science Logic (CSL 2015)*, Leibniz International Proceedings in Informatics (LIPIcs), Vol. 41, edited by S. Kreutzer (2015) pp. 211–228.

[50] A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, R. Horodecki, P. Joshi, W. Kłobus, and A. Wójcik, Physical Review Letters **112**, 120401 (2014).

[51] K. Horodecki, A. Grudka, P. Joshi, W. Kłobus, and J. Łodyga, Physical Review A **92**, 032104 (2015).

[52] B. Amaral, A. Cabello, M. T. Cunha, and L. Aolita, arXiv preprint arXiv:1705.07911 (2017).

[53] C. Duarte and B. Amaral, arXiv preprint arXiv:1711.10465 (2017).

[54] L. Henaut, L. Catani, D. E. Browne, S. Mansfield, and A. Pappa, arXiv preprint arXiv:1806.05624 (2018).

[55] S. Barz, V. Dunjko, F. Schlederer, M. Moore, E. Kashefi, and I. A. Walmsley, Physical Review A **93**, 032339 (2016).

[56] M. Clementi, A. Pappa, A. Eckstein, I. A. Walmsley, E. Kashefi, and S. Barz, arXiv preprint arXiv:1708.06144 (2017).

[57] E. F. Galvão, Physical Review A **71**, 042302 (2005).

[58] V. Veitch, C. Ferrie, D. Gross, and J. Emerson, New Journal of Physics **14**, 113011 (2012).

[59] R. W. Spekkens, Physical Review Letters **101**, 020401 (2008).

[60] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, Communications in Mathematical Physics **334**, 533 (2015).

[61] A. Cabello, S. Severini, and A. Winter, Physical Review Letters **112**, 040401 (2014).

[62] E. N. Dzhafarov, J. V. Kujala, and V. H. Cervantes, in *International Symposium on Quantum Interaction* (Springer, 2015) pp. 12–23.

[63] N. de Silva, Physical Review A **95**, 032108 (2017).

[64] J. V. Kujala, E. N. Dzhafarov, and J.-Å. Larsson, Physical Review Letters **115**, 150401 (2015).

[65] L. Wester, in Proceedings 14th International Conference on *Quantum Physics and Logic,* Nijmegen, The Netherlands, 3-7 July 2017, Electronic Proceedings in Theoretical Computer Science, Vol. 266, edited by B. Coecke and A. Kissinger (Open Publishing Association, 2018) pp. 1–22.

[66] R. Kunjwal, arXiv preprint arXiv:1709.01098 (2017).

[67] E. Amselem, L. E. Danielsen, A. J. López-Tarrida, J. R.

Portillo, M. Bourennane, and A. Cabello, Physical Review Letters **108**, 200405 (2012).

[68] S. Mansfield, "The mathematical structure of non-locality & contextuality," D.Phil. thesis, Oxford University (2013).

[69] S. Abramsky, R. S. Barbosa, and S. Mansfield, Informal Proceedings of Quantum Physics & Logic (2016).

[70] Some of these references build on earlier work relating quantum advantage to Wigner function negativity [57, 58], known to be an equivalent notion of non-classicality to contextuality [59].

[71] There now exist a number of unified treatments of non-locality and (BKS) contextuality [42, 60–62], which are closely interrelated [60, 63], and whose relations to the other kinds of non-cassicality have also been explored elsewhere [64–66].

[72] Ontological theories can be defined more generally with measures, but this will not be necessary for our present purposes.

[73] BKS contextuality has also been quantified in such a manner in [42, 50, 67–69].

[74] This is a slightly simplified version of the implementation from [21]. Note that in the homomorphism from $\mathbb{Z}_2$ to the booleans which one would perform in order to interpret the function as the logical AND gate, the roles of 0 and 1 are exchanged, i.e., $0 \mapsto 1$ and $1 \mapsto 0$.

APPENDIX

Comparison with Spekkens' contextuality

Within the ontological models framework as used by Spekkens [23], in addition to the basic ingredients of ontological models that we have set out in the main text, a number of further features are imposed or implicitly assumed motivated by the intended interpretation of ontological models. Here, we have chosen to set out a more minimal definition of what we intend to mean by ontological models, and to explicitly state any such additional assumptions as they become relevant, preferring to see these as crucial to considerations and definitions of contextuality.

Two such features are that sequential composition should be respected at the ontological level,

$$f_{U_t \cdots U_1} = f_{U_t} \circ \cdots \circ f_{U_1} \, ;$$

and that $f_{U^{(C)}} = f_{U^{(C')}}$ for sequential contexts (C) and (C'). Since these are the components of our definition of sequential transformation non-contextuality, then non-contextuality in our sense could be thought of as being implicitly baked-in to the more loaded version of the ontological models framework from the outset. From our perspective, however, this would be undesirable since it would fail to pick up on sequential transformation contextuality, a non-classical phenomenon that on the basis of our results we believe to be worthy of consideration.

Spekkens' generalised approach to non-contextuality is that ontological identifications such as $f_{U^{(C)}} = f_{U^{(C')}}$ should be imposed whenever there is an operational equivalence. Here, $U^{(C)}$ and $U^{(C')}$ would be said to be operationally equivalent if, for all choices of preparation and measurement, the outcome statistics for the prepare-transform-measure experiments with transformation $U^{(C)}$ and with transformation $U^{(C')}$ were equal. However, since $U^{(C)}$ designates the transformation U when it appears in the sequence of transformations (C), and similarly for $U^{(C')}$, such statistics are operationally inaccessible. Without broadening what it means to be operationally equivalent, our notion of non-contextuality is not captured by the Spekkens approach.

Our approach to non-contextuality, on the other hand, is to assume that operational *compositions* are respected at the ontological level, and that ontological representations are independent of operational context. Non-contextuality in the BKS and Spekkens senses are captured by this perspective as well, where now composition does not refer exclusively to sequential transformation of transformations, but also to composition of compatible measurements into a joint measurement, or composition of transformations or preparations through stochastic mixtures, etc. The presentation of the various notions of non-contextuality in the main text aims to facilitate this perspective, which will be more fully developed in a future article.

Commutativity in $\oplus L$-ontologies

In a $\oplus L$-ontology, transformations are built from CNOT and NOT gates. The action of the $\mathsf{CNOT}(i,j)$ gate with control bit i and target bit j on an ontic state $\boldsymbol{\lambda} \in \mathbb{Z}_2^s$ is a linear operation

$$\mathsf{CNOT}(i,j)\,\boldsymbol{\lambda} = (I \oplus A(j,i))\,\boldsymbol{\lambda}\,,$$

where I and $A(j,i)$ are $s \times s$ matrices over $\mathbb{Z}_2$, the former being the identity matrix and the latter the matrix whose only non-zero entry is at position (j,i). For composition of CNOT gates we have

$$\begin{aligned}\mathsf{CNOT}(k,l) \circ \mathsf{CNOT}(i,j) &= (I \oplus A(l,k))\,(I \oplus A(j,i)) \\ &= I \oplus A(l,k) \oplus A(j,i) \oplus \delta_{kj} A(l,i)\,,\end{aligned}$$

and similarly

$$\begin{aligned}\mathsf{CNOT}(i,j) \circ \mathsf{CNOT}(k,l) &= (I \oplus A(j,i))\,(I \oplus A(l,k)) \\ &= I \oplus A(j,i) \oplus A(l,k) \oplus \delta_{il} A(j,k)\,.\end{aligned}$$

The gates commute when $k \neq j, i \neq l$; i.e. the control bit for one gate cannot be the target bit for the other and vice versa.

The other basic building blocks for transformations are NOT gates. As a linear operation, the action of a NOT gate on the ith bit is simply addition by the vector $\boldsymbol{\delta}(i)$ whose only non-zero entry is in the ith position,

$$\mathsf{NOT}(i)(\boldsymbol{\lambda}) = \boldsymbol{\lambda} \oplus \boldsymbol{\delta}(i)\,.$$

NOT gates commute amongst themselves, while a NOT gate commutes with a CNOT gate if and only if it does not act on the control bit. With NOT acting on the target bit, it holds that

$$\begin{aligned}\mathsf{CNOT}(i,j)\circ\mathsf{NOT}(j)(\boldsymbol{\lambda}) &= (I\oplus A(j,i))\,(\boldsymbol{\lambda}\oplus\boldsymbol{\delta}(j))\\ &= (I\oplus A(j,i))\,\boldsymbol{\lambda}\oplus\boldsymbol{\delta}(j)\\ &= \mathsf{NOT}(j)\circ\mathsf{CNOT}(i,j)(\boldsymbol{\lambda})\,,\end{aligned}$$

where the second equality follows from the fact that $i\neq j$ since there exists a CNOT between the respective bits. With NOT acting on the control bit,

$$\begin{aligned}\mathsf{CNOT}(i,j)\circ\mathsf{NOT}(i)(\boldsymbol{\lambda}) &= (I\oplus A(j,i))\,(\boldsymbol{\lambda}\oplus\boldsymbol{\delta}(i))\\ &= (I\oplus A(j,i))\,\boldsymbol{\lambda}\oplus\boldsymbol{\delta}(i)\oplus\boldsymbol{\delta}(j)\,,\end{aligned}$$

whereas

$$\mathsf{NOT}(i)\circ\mathsf{CNOT}(i,j)(\boldsymbol{\lambda}) = (I\oplus A(j,i))\,\boldsymbol{\lambda}\oplus\boldsymbol{\delta}(i)\,.$$

In *commutative* $\oplus$*L-ontologies* we will therefore assume that the ontic state space $\mathbb{Z}_2^s$ can be partitioned into control and target bits and that NOT gates act only on target bits. This is an obvious sufficient condition for commutativity of all transformations, though weaker conditions that ensure commutativity only for subsets of the possible transformations may be interesting to consider in future work. As a linear operation, the ontological representation of any transformation U is given by

$$f_U(\boldsymbol{\lambda}) = (I\oplus A_U)\,\boldsymbol{\lambda}\oplus\boldsymbol{u}\,, \tag{11}$$

where A_U is some $s\times s$ matrix over $\mathbb{Z}_2$ containing only off-diagonal entries, $\boldsymbol{u}\in\mathbb{Z}_2^s$ gives the combined action of any NOT gates, and for composition we have

$$\begin{aligned}f_{U_t}\circ\cdots\circ f_{U_1}(\boldsymbol{\lambda}) &= \left(I\oplus\bigoplus_{i=1}^{t}A_{U_i}\right)\boldsymbol{\lambda}\oplus\bigoplus_{i=1}^{t}\boldsymbol{u_i}\\ &= \boldsymbol{\lambda}\oplus\bigoplus_{i=1}^{t}A_{U_i}\boldsymbol{\lambda}\oplus\bigoplus_{i=1}^{t}\boldsymbol{u}_i\,.\end{aligned}$$