The Contextual Fraction as a Measure of Contextuality

Samson Abramsky,[1] Rui Soares Barbosa,[1] and Shane Mansfield[2]

[1] *Department of Computer Science, University of Oxford*
Wolfson Building, Parks Rd, Oxford OX1 3QD, United Kingdom
[2] *School of Informatics, University of Edinburgh*
Informatics Forum, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom

(Dated: June 15, 2017)

We consider the contextual fraction as a quantitative measure of contextuality of empirical models, i.e. tables of probabilities of measurement outcomes in an experimental scenario. It provides a general way to compare the degree of contextuality across measurement scenarios; it bears a precise relationship to violations of Bell inequalities; its value, and a witnessing inequality, can be computed using linear programming; it is monotone with respect to the "free" operations of a resource theory for contextuality; and it measures quantifiable advantages in informatic tasks, such as games and a form of measurement based quantum computing.

PACS numbers: 03.65.Ud, 03.67.Mn

Introduction.—Recent results have established the rôle of contextuality as a resource for increasing the computational power of specific models of computation [1, 2], including enabling universal quantum computation [3]. From this perspective, it is particularly relevant to look for appropriate measures of contextuality, and indeed to pose the question of what constitutes a good measure.

In this Letter, we propose a measure of contextuality—the contextual fraction—which provides a quantitative grading between non-contextuality, at one extreme, and maximal contextuality, at the other. A maximally contextual empirical model is one that admits no proper decomposition into a convex combination of a non-contextual model and another model. In this sense, it is meaningful to consider both the non-contextual and contextual fractions of any no-signalling empirical model.

These definitions are made in the general setting of the approach to contextuality introduced in [4], in which nonlocality is seen as a special case of contextuality.

We show that the contextual fraction has a number of desirable properties: (i) it is fully general in the sense that it applies in any measurement scenario; (ii) it is bounded and normalised, taking values in the interval $[0, 1]$, with 0 indicating non-contextuality and 1 indicating strong contextuality, so it may be used to sensibly compare the degree of contextuality of empirical models not just in a given measurement scenario but also across scenarios; (iii) it has a precise relationship with violations of Bell inequalities, being the maximum normalised violation attained by the empirical model for any Bell inequality on the corresponding measurement scenario; (iv) both the contextual fraction and a witnessing Bell inequality are computable using linear programming—these were implemented and used for computational exploration of some quantum examples; (v) it is monotone with respect to a range of operations on empirical models that intuitively do not generate contextuality, and thus constitute natural "free" operations in a resource theory of contextuality, analogous to the resource theory of entanglement under LOCC operations [5], and subsuming existing resource theories for nonlocality [6–8][36]; (vi) finally, it is related to a quantifiable increase of computational power in a certain form of measurement-based quantum computation, sharpening the results of [2], and similarly to advantage in games.

We leave for future work an analysis of the relationship between the contextual fraction and other possible measures [37], and further development of (vi).

General framework for contextuality.—We briefly summarise the framework introduced in [4]. The main objects of study are **empirical models**: tables of data, specifying probability distributions over the joint outcomes of sets of compatible measurements.

A **measurement scenario** is an abstract description of a particular experimental setup. It consists of a triple $\langle X, \mathcal{M}, O\rangle$ where: X is a finite set of measurements; O is a finite set of outcome values for each measurement; and $\mathcal{M}$ is a set of subsets of X. Each $C \in \mathcal{M}$ is called a **measurement context**, and represents a set of measurements that can be performed together.

Examples of measurement scenarios include multipartite Bell-type scenarios familiar from discussions of nonlocality, Kochen–Specker configurations, measurement scenarios associated with qudit stabiliser quantum mechanics, and more. For example, the well-known $(2, 2, 2)$ Bell scenario, where two experimenters, Alice and Bob, can each choose between performing one of two different measurements, say a_1 or a_2 for Alice and b_1 or b_2 for Bob, obtaining one of two possible outcomes, is represented as follows:

$$X = \{a_1, a_2, b_1, b_2\} \qquad\qquad O = \{0, 1\}$$
$$\mathcal{M} = \{\{a_1, b_1\}, \{a_1, b_2\}, \{a_2, b_1\}, \{a_2, b_2\}\} \ .$$

Given this description of the experimental setup, then either performing repeated runs of such experiments with varying choices of measurement context and recording

TABLE I: Two empirical models on the $(2,2,2)$ Bell scenario: the well-known CHSH model [9], obtained from local projective measurements equatorial at angles 0 (for a_1, b_1) and $\pi/3$ (for a_2, b_2) on the maximally entangled two-qubit Bell-state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$; and the Popescu–Rohrlich box [10].

A	B	00	01	10	11
a_1	b_1	$1/2$	0	0	$1/2$
a_1	b_2	$3/8$	$1/8$	$1/8$	$3/8$
a_2	b_1	$3/8$	$1/8$	$1/8$	$3/8$
a_2	b_2	$1/8$	$3/8$	$3/8$	$1/8$

A	B	00	01	10	11
a_1	b_1	$1/2$	0	0	$1/2$
a_1	b_2	$1/2$	0	0	$1/2$
a_2	b_1	$1/2$	0	0	$1/2$
a_2	b_2	0	$1/2$	$1/2$	0

the frequencies of the various outcome events, or calculating theoretical predictions for the probabilities of these outcomes, results in a probability table as in Table I.

Such data is formalised as an **empirical model** for the given measurement scenario $\langle X, \mathcal{M}, O\rangle$. For each measurement context C, there is a probability distribution e_C on the joint outcomes of performing all the measurements in C; that is, on the set O^C of functions assigning an outcome in O to each measurement in C.

We require that the marginals of these distributions agree whenever contexts overlap, i.e.

$$\forall C, C' \in \mathcal{M},\ e_C|_{C\cap C'} = e_{C'}|_{C\cap C'}\ ,$$

where the notation $e_C|_U$ with $U \subseteq C$ stands for marginalisation of probability distributions (to 'forget' the outcomes of some measurements): for $t \in O^U$, $e_C|_U(t) := \sum_{s\in O^C, s|_U = t} e_C(s)$. The requirement of **compatibility of marginals** is a generalisation of the usual **no-signalling** condition, and is satisfied in particular by all empirical models arising from quantum predictions [4].

An empirical model is said to be **contextual** if this family of distributions cannot itself be obtained as the marginals of a single probability distribution on global assignments of outcomes to all measurements, i.e. a distribution d on O^X (where O^X acts as a canonical set of deterministic hidden variables) such that $\forall C \in \mathcal{M},\ d|_C = e_C$. Equivalently [4], contextual empirical models are those which have no realisation by factorisable hidden variable models; thus for Bell-type measurement scenarios contextuality specialises to the usual notion of **non-locality**.

In certain cases, one can witness contextuality from merely the possibilistic, rather than probabilistic, information contained in an empirical model—i.e. which events are possible (with non-zero probability) and which are impossible (with zero probability). A yet stronger form of contextuality occurs when no global assignment of outcomes is even consistent with the possible events: an empirical model e is said to be **strongly contextual** if there is no global assignment $g \in O^X$ such that $\forall C \in \mathcal{M},\ e_C(g|_C) > 0$. An example is the Popescu–Rohrlich box (Table I). This is the highest level in the qualitative hierarchy of strengths of contextuality introduced in [4].

The contextual fraction.—Given two empirical models e and e' on the same measurement scenario and $\lambda \in [0,1]$, we define the empirical model $\lambda e + (1-\lambda)e'$ by taking the convex sum of probability distributions at each context. Compatibility is preserved by this convex sum, hence it yields a well-defined empirical model.

A natural question to ask is: what fraction of a given empirical model e admits a non-contextual explanation? This approach enables a refinement of the binary notion of contextuality *vs* non-contextuality into a quantitative grading. Instead of asking for a probability distribution on global assignments that marginalises to the empirical distributions at each context, we ask only for a subprobability distribution [38] b on global assignments O^X that marginalises at each context to a subdistribution of the empirical data, thus explaining a fraction of the events, i.e. $\forall C \in \mathcal{M},\ b|_C \leq e_C$. Equivalently, we ask for a convex decomposition

$$e = \lambda e^{NC} + (1-\lambda)e' \tag{1}$$

where e^{NC} is a non-contextual model and e' is another (no-signalling) empirical model. The maximum weight of such a global subprobability distribution, or the maximum possible value of λ in such a decomposition is called the **non-contextual fraction** of e, and generalises the **local fraction** previously introduced for models on Bell-type scenarios [11] [39]. We denote it by $\mathsf{NCF}(e)$, and the contextual fraction by $\mathsf{CF}(e) := 1 - \mathsf{NCF}(e)$.

The notion of contextual fraction in general scenarios was introduced in [4], where it was proved that a model is strongly contextual if and only if its contextual fraction is 1. In fact, in any convex decomposition of the form (1) giving maximal weight to the non-contextual model, the other model will necessarily be strongly contextual. This means that any empirical model e admits a convex decomposition

$$e = \mathsf{NCF}(e)\, e^{NC} + \mathsf{CF}(e)\, e^{SC} \tag{2}$$

into a non-contextual and a strongly contextual model. Note that e^{NC} and e^{SC} are not necessarily unique.

Computing the contextual fraction via LP.—The task of finding a consistent probability subdistribution with maximum weight for a given empirical model can be formulated as a linear programming problem. This is a relaxation of the test for contextuality by solving a system of linear equations over the nonnegative reals from [4].

Fix a measurement scenario $\langle X, \mathcal{M}, O\rangle$. Let $n := |O^X|$ be the number of global assignments g, and $m := \sum_{C\in\mathcal{M}} |O^C| = |\{\langle C, s\rangle \mid C \in \mathcal{M}, s \in O^C\}|$ be the number of local assignments ranging over contexts. The **incidence matrix** [4] $\mathbf{M}$ is an $m \times n$ $(0,1)$-matrix that records the restriction relation between global and local assignments:

$$\mathbf{M}[\langle C, s\rangle, g] := \begin{cases} 1 & \text{if } g|_C = s; \\ 0 & \text{otherwise.} \end{cases}$$

An empirical model e can be represented as a vector $\mathbf{v}^e \in \mathbb{R}^m$, with the component $\mathbf{v}^e[\langle C, s\rangle]$ recording the probability given by the model to the assignment s at the measurement context C, $e_C(s)$. This vector is a flattened version of the table used to represent the empirical model (e.g. Table I). The columns of the incidence matrix, $\mathbf{M}[-, g]$, are the vectors corresponding to the (non-contextual) deterministic models obtained from global assignments $g \in O^X$. Recall that every non-contextual model can be written as a mixture of these. A probability distribution on global assignments can be represented as a vector $\mathbf{d} \in \mathbb{R}^n$ with non-negative components summing to 1, and then the corresponding non-contextual model is represented by the vector $\mathbf{M}\,\mathbf{d}$. So a model e is non-contextual if and only if there exists a $\mathbf{d} \in \mathbb{R}^n$ such that:

$$\mathbf{M}\,\mathbf{d} \;=\; \mathbf{v}^e \qquad \text{and} \qquad \mathbf{d} \geq \mathbf{0}\;.$$

Note that the first condition implies that $\mathbf{d}$ is normalised.

A global subprobability distribution is also represented by a vector $\mathbf{b} \in \mathbb{R}^n$ with non-negative components, its weight being given by the dot product $\mathbf{1} \cdot \mathbf{b}$, where $\mathbf{1} \in \mathbb{R}^n$ is the vector whose n components are each 1. The following LP thus calculates the non-contextual fraction of an empirical model e, with $\mathsf{NCF}(e) = \mathbf{1} \cdot \mathbf{b}^*$ where $\mathbf{b}^*$ is an optimal solution:

$$\begin{array}{ll} \texttt{Find} & \mathbf{b} \in \mathbb{R}^n \\ \texttt{maximising} & \mathbf{1} \cdot \mathbf{b} \\ \texttt{subject to} & \mathbf{M}\,\mathbf{b} \;\leq\; \mathbf{v}^e \\ \texttt{and} & \mathbf{b} \;\geq\; \mathbf{0} \quad . \end{array} \tag{3}$$

Violations of generalised Bell inequalities.—We now provide further justification for viewing the contextual fraction as a measure of contextuality by relating it to violations of contextuality-witnessing inequalities.

An **inequality** for a scenario $\langle X, \mathcal{M}, O\rangle$ is given by a vector $\mathbf{a} \in \mathbb{R}^m$ of real coefficients indexed by local assignments $\langle C, s\rangle$, and a bound R. For a model e, the inequality reads $\mathbf{a} \cdot \mathbf{v}^e \;\leq\; R$, where

$$\mathbf{a} \cdot \mathbf{v}^e \;=\; \sum_{C \in \mathcal{M}, s \in O^C} \mathbf{a}[\langle C, s\rangle]\, e_C(s)\;.$$

Without loss of generality, we can take R to be non-negative (in fact, even $R = 0$) as any inequality is equivalent to one of this form. We call it a **Bell inequality** if it is satisfied by every non-contextual model. This generalises the usual notion of Bell inequality, which is defined for Bell-type scenarios for nonlocality, to apply to any contextuality scenario. If, moreover, it is saturated by some non-contextual model, the Bell inequality is said to be **tight**. A Bell inequality establishes a bound for the value of $\mathbf{a} \cdot \mathbf{v}^e$ amongst non-contextual models e. For more general models, this quantity is limited only by the algebraic bound

$$\|\mathbf{a}\| := \sum_{C \in \mathcal{M}} \max\left\{\mathbf{a}[\langle C, s\rangle] \mid s \in O^C\right\}\;.$$

Note that we will consider only inequalities satisfying $R < \|\mathbf{a}\|$, which excludes inequalities trivially satisfied by all models, and avoids cluttering the presentation with special caveats about division by 0.

The **violation** of a Bell inequality $\langle \mathbf{a}, R\rangle$ by a model e is $\max\{0, \mathbf{a} \cdot \mathbf{v}^e - R\}$. However, it is useful to normalise this value by the maximum possible violation in order to give a better idea of the *extent* to which the model violates the inequality. The **normalised violation** of the Bell inequality by the model e is

$$\frac{\max\{0, \mathbf{a} \cdot \mathbf{v}^e - R\}}{\|\mathbf{a}\| - R}\;.$$

Theorem 1. *Let e be an empirical model. (i) The normalised violation by e of any Bell inequality is at most $\mathsf{CF}(e)$; (ii) if $\mathsf{CF}(e) > 0$, this bound is attained, i.e. there exists a Bell inequality whose normalised violation by e is $\mathsf{CF}(e)$; (iii) moreover, for any decomposition of the form* (2)*, this Bell inequality is tight at the non-contextual model e^{NC} (provided $\mathsf{NCF}(e) > 0$) and maximally violated at the strongly contextual model e^{SC}.*

The proof of this result is based on the Strong Duality theorem of linear programming [12]. It provides an LP method of calculating a witnessing Bell inequality for any empirical model e. The symmetric dual of (3) is the following LP:

$$\begin{array}{ll} \texttt{Find} & \mathbf{y} \in \mathbb{R}^m \\ \texttt{minimising} & \mathbf{y} \cdot \mathbf{v}^e \\ \texttt{subject to} & \mathbf{M}^T\,\mathbf{y} \;\geq\; \mathbf{1} \\ \texttt{and} & \mathbf{y} \;\geq\; \mathbf{0} \quad . \end{array} \tag{4}$$

The Strong Duality theorem says that, if $\mathbf{b}^*$ is a solution for (3), then there is a solution $\mathbf{y}^*$ for (4) satisfying

$$\mathbf{1} \cdot \mathbf{b}^* \;=\; \mathbf{y}^* \cdot \mathbf{v}^e. \tag{5}$$

Defining $\mathbf{a}^* \;:=\; |\mathcal{M}|^{-1}\mathbf{1} - \mathbf{y}^*$, one can show using (5) that the Bell inequality determined by $\mathbf{a}^*$ as the vector of coefficients and with bound $R = 0$ satisfies parts (ii) and (iii) of the Theorem. A detailed proof is provided in the supplemental material [40].

Monotonicity.—A key desideratum of a useful measure of contextuality is that it be a monotone for the free operations of a resource theory for contextuality. A fuller treatment of this subject will be presented in a forthcoming article by the authors; here, we consider the properties of the contextual fraction with respect to some of these operations.

We consider the following operations: first, translation of measurements (including restriction to a smaller set of measurements, replication of measurements, etc.); secondly, coarse-graining of outcomes. Special cases of these give isomorphic relabelling of measurements and outcomes. We also consider operations that combine two empirical models to build a new one. The first of these is **probabilistic mixing** with a weight $\lambda \in [0,1]$. The second is **controlled choice**: given empirical models e and e' on scenarios $\langle X, \mathcal{M}, O\rangle$ and $\langle X', \mathcal{M}', O\rangle$ respectively, $e \,\&\, e'$ is defined on the scenario $\langle X \sqcup X', \mathcal{M} \sqcup \mathcal{M}', O\rangle$ by $(e \,\&\, e')_C := e_C$ for $C \in \mathcal{M}$ and $(e \,\&\, e')_{C'} := e_{C'}$ for $C' \in \mathcal{M}'$. The third is a **product**: $e \otimes e'$ is an empirical model defined on the scenario $\langle X \sqcup X', \mathcal{M} \star \mathcal{M}', O\rangle$, where $\mathcal{M} \star \mathcal{M}' := \{C \sqcup C' \mid C \in \mathcal{M}, C' \in \mathcal{M}'\}$, by $(e \otimes e')_{C \sqcup C'}\langle s, s'\rangle := e_C(s)\, e'_{C'}(s')$ for all $C \in M$, $C' \in \mathcal{M}'$, $s \in O^C$, and $s' \in O^{C'}$.

These operations can be used to construct any local empirical model on Bell scenarios starting from a very simple "generator": a deterministic model over a single measurement. This is illustrated in the supplemental material.

Theorem 2. *The contextual fraction is invariant under relabellings, and non-increasing under translation of measurements and coarse-graining of outcomes. For the combining operations, it satisfies the following properties:*

- $\mathsf{CF}(\lambda e + (1-\lambda)e') \leq \lambda \mathsf{CF}(e) + (1-\lambda)\mathsf{CF}(e)$
- $\mathsf{CF}(e \,\&\, e') = \max\{\mathsf{CF}(e), \mathsf{CF}(e')\}$
- $\mathsf{CF}(e \otimes e') = \mathsf{CF}(e) + \mathsf{CF}(e') - \mathsf{CF}(e)\mathsf{CF}(e')$

A consequence of this result is that, for any of the combining operations, when e' is a non-contextual model (and thus composing with e' is a free operation), CF acts as a monotone: the contextual fraction of the new model is at most that of e (in fact, with equality holding for both choice and product).

Computational explorations.—General computational tools in the form of a *Mathematica* package have been developed implementing the two LPs described above to calculate, for *any* empirical model in *any* scenario: the (non-)contextual fraction, a decomposition of the form (2), and the generalised Bell inequality from Theorem 1-(ii) for which the maximal violation is achieved. The package also calculates quantum empirical models from any (pure or mixed) state and any specified sets of compatible measurements.

As an example to illustrate the use of this package, we consider the empirical models obtained from local measurements on various n-qubit states. On each qubit, we allow the same two local measurements, equatorial on the Bloch sphere, parametrised by angles ϕ_1 and ϕ_2. Figure 1 plots the contextual fraction of the resulting models as a function of these angles.

Computational explorations of this kind can be a useful tool for guiding research, pointing the way to conjectures and results (e.g. [13–15]). A more detailed analysis of the examples from Figure 1, leading to the characterisation of a family of strongly contextual models arising from n-partite GHZ states, can be found in the supplemental material.

Applications to quantum computation.—Contextuality has been associated with quantum advantage in certain information-processing and computational tasks. One use for a measure of contextuality is to quantify such advantages.

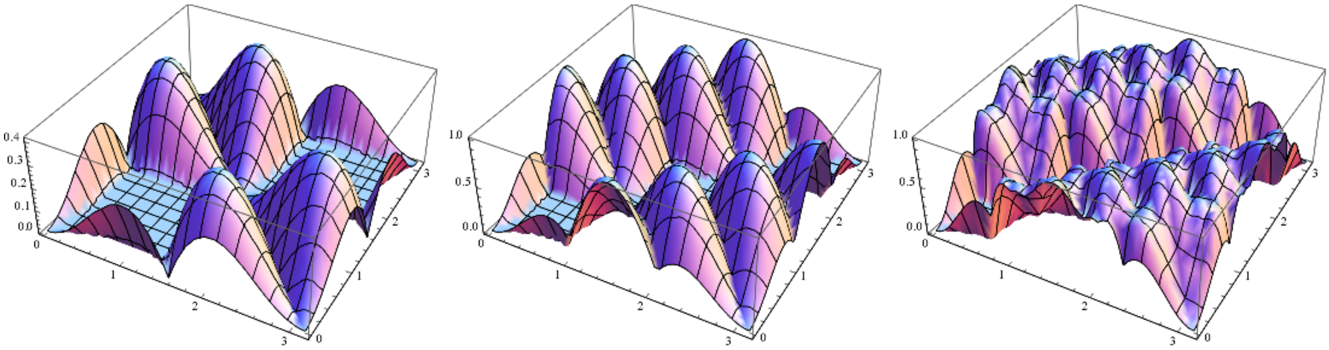
One computational model in which contextuality has been associated with an advantage is measurement-based quantum computation (MBQC). An $l2$-MBQC is a process with m classical bits of input and l of output, using an $(n,2,2)$ empirical model (n parties, 2 measurement settings per party, 2 outcomes per measurement) as a resource. The classical control—which pre-processes the inputs, determines the flow of measurements by choosing which sites to measure next and with which measurement setting (potentially depending on previous outcomes), and post-processes to produce the outputs—can only perform $\mathbb{Z}_2$-linear computations. The additional power to compute non-linear functions thus resides in certain resource empirical models.

In [2, Theorem 2] it was shown that if an $l2$-MBQC process *deterministically* calculates a non-$\mathbb{Z}_2$-linear Boolean function $f : 2^m \longrightarrow 2^l$, then the resource is necessarily strongly contextual. A probabilistic version was also obtained in [2, Theorem 3]: contextuality must be present whenever a non-linear function is calculated with a sufficiently large probability of success. By analysing that proof, we extract a sharpened version of this result establishing a precise relationship between the hardness (non-linearity) of the function, the probability of success, and the contextual fraction.

The **average distance** between two Boolean functions $f, g : 2^m \longrightarrow 2^l$ is given by $\tilde{d}(f,g) := 2^{-m}|\{\mathbf{i} \in 2^m \mid f(\mathbf{i}) \neq g(\mathbf{i})\}|$. The average distance of f to the closest $\mathbb{Z}_2$-linear function is denoted by $\tilde{\nu}(f)$.

Theorem 3. *Let $f : 2^m \longrightarrow 2^l$ be a Boolean function and consider an $l2$-MBQC that uses the empirical model e to compute f with average success probability $\bar{p}_S$ over all 2^m possible inputs, and corresponding average failure probability $\bar{p}_F = 1 - \bar{p}_S$. Then, $\bar{p}_F \geq \mathsf{NCF}(e)\tilde{\nu}(f)$.*

Note that for deterministic computation ($\bar{p}_S = 1$) of a non-linear function ($\tilde{\nu} > 0$), we require strong contextuality ($\mathsf{NCF}(e) = 0$), recovering the deterministic result in [2]. More generally, for a given non-linear function, the higher the desired success probability the larger the contextual fraction must be. Additional details, including a rigorous presentation and proof, may be found in the supplemental material.



(a) (b) (c)

FIG. 1: Plots of the contextual fraction for empirical models obtained with projective measurements at ϕ_1 and ϕ_2 in the X–Y plane for each qubit on the states: (a) the Bell state $|\Phi^+\rangle := \frac{|00\rangle+|11\rangle}{\sqrt{2}}$; (b) $|\psi_{\mathsf{GHZ}(3)}\rangle$; (c) $|\psi_{\mathsf{GHZ}(4)}\rangle$, where the n-partite GHZ state ($n > 2$) is given by $|\psi_{\mathsf{GHZ}(n)}\rangle = \frac{|0\rangle^{\otimes n}+|1\rangle^{\otimes n}}{\sqrt{2}}$.

Similar results can be obtained to quantify advantage in games, generalizing XOR games on Bell scenarios [16]. A game is specified by n Boolean formulae, one for each context, which describe the winning condition that the output must satisfy. If the formulae are k-consistent, meaning that at most k of them have a joint satisfying assignment, then the *hardness* of the game is measured by $\frac{(n-k)}{n}$. One can show that $\bar{p}_F \geq \mathsf{NCF}(e)\frac{(n-k)}{n}$, relating the probability of success, the non-contextual fraction, and the hardness of the task. See [17] for the relation with Bell inequalities, from which a proof of this result follows. Details are given in the supplemental material, Theorem 4. Further development of these ideas is a topic for future research.

Acknowledgements.—This work was done in part while the authors visited the Simons Institute for the Theory of Computing (supported by the Simons Foundation) at the University of California, Berkeley, as participants of the Logical Structures in Computation programme, and while SM was based at l'Institut de Recherche en Informatique Fondamentale, Université Paris Diderot – Paris 7. Support from the following is also gratefully acknowledged: EPSRC EP/N018745/1, Contextuality as a Resource in Quantum Computation (SA, RSB); Fondation Sciences Mathématiques de Paris, postdoctoral research grant eotpFIELD15RPOMT-FSMP1, Contextual Semantics for Quantum Theory (SM); the Oxford Martin School James Martin Program on Bioinspired Quantum Technologies, (SA, SM, RSB); FCT – Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology), PhD grant SFRH/BD/94945/2013 (RSB).

[1] J. Anders and D. E. Browne, Phys. Rev. Lett. **102**, 050502 (2009).
[2] R. Raussendorf, Phys. Rev. A **88**, 022322 (2013).
[3] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Nature **510**, 351 (2014).
[4] S. Abramsky and A. Brandenburger, New J. Phys. **13**, 113036 (2011).
[5] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009).
[6] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Phys. Rev. A **71**, 022101 (2005).
[7] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, Phys. Rev. Lett. **109**, 070401 (2012).
[8] J. I. de Vicente, J. Phys. A: Math. Theor. **47**, 424017 (2014).
[9] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[10] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[11] A. C. Elitzur, S. Popescu, and D. Rohrlich, Phys. Lett. A **162**, 25 (1992).
[12] G. B. Dantzig and M. N. Thapa, **Linear programming 2: Theory and extenstions**, Springer Series in Operations Research and Financial Engineering (Springer Verlag, 2003).
[13] S. Abramsky, C. M. Constantin, and S. Ying, Inform. Comput. **250**, 3 (2016).
[14] S. Abramsky and C. Constantin, Electronic Proceedings in Theoretical Computer Science **171**, 10 (2014).
[15] S. Mansfield, DPhil thesis, University of Oxford (2013).
[16] R. Cleve, P. Høyer, B. Toner, and J. Watrous, in **Proceedings of 19th IEEE Annual Conference on Computational Complexity (CCC 2004)** (2004), pp. 236–249.
[17] S. Abramsky and L. Hardy, Phys. Rev. A **85**, 062114 (2012).
[18] B. Amaral, A. Cabello, M. Terra Cunha, and L. Aolita, **Noncontextual wirings**, available as arXiv:1705.07911 [quant-ph].
[19] E. N. Dzhafarov, J. V. Kujala, and V. H. Cervantes,

in **International Symposium on Quantum Interaction** (Springer, 2015), pp. 12–23.

[20] D. Kaszlikowski, P. Gnaciński, M. Żukowski, W. Miklaszewski, and A. Zeilinger, Phys. Rev. Lett. **85**, 4418 (2000).

[21] A. Acín, T. Durt, N. Gisin, and J. I. Latorre, Phys. Rev. A **65**, 052325 (2002).

[22] M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf, Phys. Rev. Lett. **104**, 170405 (2010).

[23] S. Massar, S. Pironio, J. Roland, and B. Gisin, Phys. Rev. A **66**, 052112 (2002).

[24] S. Laplante, V. Lerays, and J. Roland, in **Automata, Languages, and Programming: 39th International Colloquium (ICALP 2012), Proceedings, Part I**, edited by A. Czumaj, K. Mehlhorn, A. Pitts, and R. Wattenhofer (Springer, 2012), vol. 7391 of **Lecture Notes in Computer Science**, pp. 617–628.

[25] J. Degorre, M. Kaplan, S. Laplante, and J. Roland, in **Mathematical Foundations of Computer Science 2009: 34th International Symposium (MFCS 2009), Proceedings**, edited by R. Královič and D. Niwiński (Springer, 2009), vol. 5734 of **Lecture Notes in Computer Science**, pp. 270–281.

[26] S. Laplante, M. Laurière, A. Nolin, J. Roland, and G. Senno, in **11th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2016)**, edited by A. Broadbent (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2016), vol. 61 of **Leibniz International Proceedings in Informatics (LIPIcs)**, pp. 5:1–5:24.

[27] J. Barrett, A. Kent, and S. Pironio, Phys. Rev. Lett. **97**, 170409 (2006).

[28] L. Aolita, R. Gallego, A. Acín, A. Chiuri, G. Vallone, P. Mataloni, and A. Cabello, Phys. Rev. A **85**, 032107 (2012).

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

[30] N. D. Mermin, Phys. Rev. Lett. **65**, 1838 (1990).

[31] S. Abramsky, S. Mansfield, and R. S. Barbosa, in **Proceedings 8th International Workshop on Quantum Physics and Logic, Nijmegen, Netherlands, October 27–29, 2011**, edited by B. Jacobs, P. Selinger, and B. Spitters (Open Publishing Association, 2012), vol. 95 of **Electronic Proceedings in Theoretical Computer Science**, pp. 1–14.

[32] R. Cleve and R. Mittal, in **Automata, Languages, and Programming: 41st International Colloquium (ICALP 2014), Proceedings, Part I**, edited by J. Esparza, P. Fraigniaud, T. Husfeldt, and E. Koutsoupias (Springer, 2014), vol. 8572 of **Lecture Notes in Computer Science**, pp. 320–331.

[33] L. Mančinska and D. E. Roberson, J. Comb. Theory B **118**, 228 (2016).

[34] R. Cleve, L. Liu, and W. Slofstra, J. Math. Phys. **58**, 012202 (2017).

[35] S. Abramsky, R. S. Barbosa, N. de Silva, and O. Zapata, **The quantum monad on relational structures**, Submitted for publication, available as arXiv:1705.07310 [cs.LO] (2017).

[36] See also the recent [18], which develops a different approach to a resource theory for contextuality.

[37] These include a negative probability measure introduced in [4], contextuality-by-default measures [19], and various other measures considered for nonlocality in specific measurement scenarios, e.g. noise-based [20–22] and inefficiency-based [23–25] measures, known to relate to communication complexity [25, 26].

[38] A subprobability distribution on a set S is a map $b : S \longrightarrow \mathbb{R}_{\geq 0}$ with finite support and $w(b) \leq 1$, where $w(b) := \sum_{s \in S} b(s)$ is called its weight. The set of subprobability distributions on S is ordered pointwise: b' is a subdistribution of b (written $b' \leq b$) whenever $\forall s \in S,\ b'(s) \leq b(s)$.

[39] See also [27, 28] where the term *local fraction* is actually used.

[40] See Supplemental Material at [URL will be inserted by publisher] for rigorous proofs and further details on our computational explorations.

The Contextual Fraction as a Measure of Contextuality—Supplemental Material

Samson Abramsky,[1] Rui Soares Barbosa,[1] and Shane Mansfield[2]

[1]*Department of Computer Science, University of Oxford*
Wolfson Building, Parks Rd, Oxford OX1 3QD, United Kingdom
[2]*School of Informatics, University of Edinburgh*
Informatics Forum, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom

(Dated: June 15, 2017)

PACS numbers: 03.65.Ud, 03.67.Mn

A. Contextual fraction and violations of Bell inequalities (Proof of Theorem 1)

Theorem 1 establishes the close link between the contextual fraction of an empirical model and the violation of generalised Bell inequalities by that model. In the main text, we presented this result along with the broad idea of its proof, with particular emphasis on the linear program that calculates the Bell inequality for which a given model achieves a maximal violation. For completeness, we include here the detailed proof of the three statements.

As a preliminary remark, we note that the Strong Duality theorem of Linear Programming does indeed apply here, since the set of feasible solutions is bounded and non-empty. Indeed, non-emptiness holds since the zero vector is feasible, while boundedness holds since the constraints imply that every component of a vector satisfying them must lie in the unit interval.

Theorem 1. *Let e be an empirical model. (i) The normalised violation by e of any Bell inequality is at most $\mathsf{CF}(e)$; (ii) if $\mathsf{CF}(e) > 0$, this bound is attained, i.e. there exists a Bell inequality whose normalised violation by e is $\mathsf{CF}(e)$; (iii) moreover, for any decomposition of the form $e = \mathsf{NCF}(e)e^{NC} + \mathsf{CF}(e)e^{SC}$, this Bell inequality is tight at the non-contextual model e^{NC} (provided $\mathsf{NCF}(e) > 0$) and maximally violated at the strongly contextual model e^{SC}.*

Proof. **(i)** This follows from the decomposition of e into a non-contextual and a strongly contextual models,

$$e = \mathsf{NCF}(e)e^{NC} + \mathsf{CF}(e)e^{SC} \ .$$

For any Bell inequality, determined by $\langle \mathbf{a}, R\rangle$, with $\|\mathbf{a}\| > R$, the left-hand side of the inequality for the model e adds up to

$$\begin{aligned}
& \mathbf{a}\cdot\mathbf{v}^e \\
= \quad & \{ \text{ by decomposition above and linearity of } (\mathbf{a}\cdot -) \ \} \\
& \mathsf{NCF}(e)\,\mathbf{a}\cdot\mathbf{v}^{e^{NC}} \ + \ \mathsf{CF}(e)\,\mathbf{a}\cdot\mathbf{v}^{e^{SC}} \\
\leq \quad & \{ \text{ since } \mathbf{a}\cdot\mathbf{v}^{e^{NC}} \leq R \text{ and } \mathbf{a}\cdot\mathbf{v}^{e^{SC}} \leq \|\mathbf{a}\| \ \} \\
& \mathsf{NCF}(e)\,R \ + \ \mathsf{CF}(e)\,\|\mathbf{a}\| \\
= \quad & \\
& \mathsf{NCF}(e)\,R \ + \ \mathsf{CF}(e)\,R \ + \ \mathsf{CF}(e)\,(\|\mathbf{a}\| - R) \\
= \quad & \{ \text{ by } \mathsf{NCF}(e) + \mathsf{CF}(e) = 1 \ \} \\
& R \ + \ \mathsf{CF}(e)\,(\|\mathbf{a}\| - R) \ .
\end{aligned}$$

Therefore, since $\|\mathbf{a}\| > R$, we have, as required,

$$\frac{\mathbf{a}\cdot\mathbf{v}^e - R}{\|\mathbf{a}\| - R} \ \leq \ \mathsf{CF}(e) \ .$$

(ii) Recall the LP (3) that yields the (non-)contextual fraction of a model e. Its symmetric dual is the following:

$$\begin{array}{ll}
\texttt{Find} & \mathbf{y} \in \mathbb{R}^m \\
\texttt{minimising} & \mathbf{v}^e \cdot \mathbf{y} \\
\texttt{subject to} & \mathbf{M}^\mathsf{T}\,\mathbf{y} \geq \mathbf{1} \\
\texttt{and} & \mathbf{y} \geq \mathbf{0} \quad .
\end{array}$$

Under the transformation of variables

$$\mathbf{a} \ := \ |\mathcal{M}|^{-1}\mathbf{1} \ - \ \mathbf{y} \ , \tag{6}$$

where $|\mathcal{M}|$ is the number of maximal contexts in the scenario, the above LP can then be equivalently restated as follows:

$$\begin{array}{ll}
\texttt{Find} & \mathbf{a} \in \mathbb{R}^m \\
\texttt{maximising} & \mathbf{a}\cdot\mathbf{v}^e \\
\texttt{subject to} & \mathbf{M}^\mathsf{T}\,\mathbf{a} \leq \mathbf{0} \\
\texttt{and} & \mathbf{a} \leq |\mathcal{M}|^{-1}\mathbf{1} \ ,
\end{array} \tag{7}$$

with solutions (resp. optimal solutions) of one optimisation problem corresponding bijectively to those of the other via the transformation (6). To see that this is the case, note that:

$$\begin{aligned}
& \mathbf{M}^\mathsf{T}\,\mathbf{a} \ \leq \ \mathbf{0} \\
\Leftrightarrow \quad & \{ \text{ by eq. (6) } \} \\
& \mathbf{M}^\mathsf{T}\,(|\mathcal{M}|^{-1}\mathbf{1} \ - \ \mathbf{y}) \ \leq \ \mathbf{0} \\
\Leftrightarrow \quad & \{ \text{ by linearity } \}
\end{aligned}$$

$$
\begin{aligned}
& |\mathcal{M}|^{-1}\mathbf{M}^{\mathsf{T}}\,\mathbf{1} - \mathbf{M}^{\mathsf{T}}\,\mathbf{y} \;\leq\; \mathbf{0} \\
\Leftrightarrow \quad & \{ \text{ } \mathbf{M}^{\mathsf{T}} \text{ has exactly } |\mathcal{M}| \text{ 1-entries in each row } \} \\
& |\mathcal{M}|^{-1}|\mathcal{M}|\mathbf{1} - \mathbf{M}^{\mathsf{T}}\,\mathbf{y} \;\leq\; \mathbf{0} \\
\Leftrightarrow \quad & \\
& \mathbf{M}^{\mathsf{T}}\,\mathbf{y} \;\geq\; \mathbf{1}
\end{aligned}
$$

and

$$
\begin{aligned}
& \mathbf{a} \;\leq\; |\mathcal{M}|^{-1}\mathbf{1} \\
\Leftrightarrow \quad & \{ \text{ by eq. (6) } \} \\
& |\mathcal{M}|^{-1}\mathbf{1} - \mathbf{y} \;\leq\; |\mathcal{M}|^{-1}\mathbf{1} \\
\Leftrightarrow \quad & \\
& \mathbf{y} \;\geq\; \mathbf{0} \;,
\end{aligned}
$$

showing that the feasibility conditions of the LPs (4) and (7) are equivalent, and moreover that

$$
\begin{aligned}
& \mathbf{a} \cdot \mathbf{v}^e \\
= \quad & \{ \text{ by eq. (6) } \} \\
& \big(|\mathcal{M}|^{-1}\mathbf{1} - \mathbf{y}\big) \cdot \mathbf{v}^e \\
= \quad & \{ \text{ by linearity } \} \\
& |\mathcal{M}|^{-1}\mathbf{1} \cdot \mathbf{v}^e - \mathbf{y} \cdot \mathbf{v}^e \\
= \quad & \{ \text{ } \mathbf{v}^e \text{ consists of } |M| \text{ probability distributions } \} \\
& 1 - \mathbf{y} \cdot \mathbf{v}^e \;, \qquad (8)
\end{aligned}
$$

showing that $\mathbf{y} \cdot \mathbf{v}^e$ is minimised exactly when $\mathbf{a} \cdot \mathbf{v}^e$ is maximised.

The idea is that the components of a solution vector $\mathbf{a}$ (indexed by local assignments $\langle C, s \rangle$) are to be taken as the coefficients of an inequality, with bound $R = 0$.

We first show that any feasible solution of the LP determines a Bell inequality, i.e. an inequality that is satisfied by all non-contextual models. It suffices to show that it is satisfied by the deterministic non-contextual models—that is, those determined by a single global assignment $g : X \longrightarrow O$—since all other non-contextual models are convex combinations of these. Recall that the columns of $\mathbf{M}$ (and so the rows of $\mathbf{M}^{\mathsf{T}}$) are exactly the vectors representing these models. Hence, the fact that $\mathbf{a}$ determines a Bell inequality is concisely expressed by the system of linear inequalities

$$
\mathbf{M}^{\mathsf{T}}\,\mathbf{a} \;\leq\; \mathbf{0} \;,
$$

which is one of the feasibility conditions of our LP (7).

The other feasibility condition, $\mathbf{a} \;\leq\; |\mathcal{M}|^{-1}\mathbf{1}$, is a bound on the components of $\mathbf{a}$. This acts as a normalisation condition guaranteeing that, for any feasible solution, the algebraic bound of the inequality (i.e. its maximal violation) is at most 1:

$$
\begin{aligned}
\|\mathbf{a}\| \;&=\; \sum_{C \in \mathcal{M}} \max \big\{ \mathbf{a}[\langle C, s \rangle] \mid s \in O^C \big\} \\
&\leq\; \sum_{C \in \mathcal{M}} |\mathcal{M}|^{-1} \\
&=\; 1 \;.
\end{aligned}
$$

Consequently,

$$
\max\{0, \mathbf{a} \cdot \mathbf{v}^e\} \;\;\geq\;\; (\mathbf{a} \cdot \mathbf{v}^e)\; \max\{0, \|\mathbf{a}\|\} \;,
$$

and so (whenever the inequality corresponding to $\mathbf{a}$ is of any interest, i.e. whenever $0 < \|\mathbf{a}\|$) the normalised violation of the inequality by e is at least $\mathbf{a} \cdot \mathbf{v}^e$, the objective function that the LP maximises. (In fact, for an optimal solution, the two bounds above are attained: provided e is contextual, the algebraic bound of the optimal inequality is 1 and its normalised violation by e is simply given by $\mathbf{a} \cdot \mathbf{v}^e$, as we shall see at the end of this item.)

We now show that an optimal solution, $\mathbf{a}^*$, to the LP yields our desired inequality, whose violation by e is $\mathsf{CF}(e)$. Let $\mathbf{b}^*$ denote an optimal solution to the primal LP (3), meaning that $\mathbf{1} \cdot \mathbf{b}^* = \mathsf{NCF}(e)$. The Strong Duality theorem of Linear Programming (see e.g. [12]) then says that the dual LP (4) also admits an optimal solution $\mathbf{y}^*$ and moreover that these two optimal solutions are related by

$$
\mathbf{y}^* \cdot \mathbf{v}^e \;=\; \mathbf{1} \cdot \mathbf{b}^* \;. \qquad (9)
$$

Hence, writing $\mathbf{a}^*$ for the corresponding (via eq. (6)) optimal solution of (7), we have

$$
\begin{aligned}
& \mathbf{a}^* \cdot \mathbf{v}^e \\
= \quad & \{ \text{ by eq. (8) } \} \\
& 1 - \mathbf{y}^* \cdot \mathbf{v}^e \\
= \quad & \{ \text{ by eq. (9) (strong duality) } \} \\
& 1 - \mathbf{1} \cdot \mathbf{b}^* \\
= \quad & \{ \text{ } \mathbf{b}^* \text{ is optimal solution to LP (3) } \} \\
& 1 - \mathsf{NCF}(e) \\
= \quad & \\
& \mathsf{CF}(e)
\end{aligned}
$$

This shows that the normalised violation of the inequality by the model e is at least $\mathsf{CF}(e)$. Since the opposite inequality follows from the first item, this concludes the proof that the model e attains a normalised violation of $\mathsf{CF}(e)$ of the Bell inequality with coefficients $\mathbf{a}^*$ and bound 0. Incidentally, this implies in particular that the algebraic bound of this inequality, $\|\mathbf{a}^*\|$, is equal to 1.

(iii) Consider any decomposition of the model e (which satisfies $\mathsf{CF}(e) > 0$) into a non-contextual and a strongly contextual parts, with maximal possible weight on the former:

$$
e \;=\; \mathsf{NCF}(e)\, e^{NC} \;+\; \mathsf{CF}(e)\, e^{SC} \;. \qquad (10)
$$

We want to show that the non-contextual part of the model, e^{NC}, saturates the inequality from the previous item, and that the strongly contextual part, e^{SC}, maximally violates it (i.e. achieves the algebraic bound, which we know from the previous item to be 1). That is, the goal is to show that

$$\mathbf{a}^* \cdot \mathbf{v}^{e^{NC}} = 0 \qquad \text{and} \qquad \mathbf{a}^* \cdot \mathbf{v}^{e^{SC}} = 1 \ .$$

Note that the inequalities

$$\mathbf{a}^* \cdot \mathbf{v}^{e^{NC}} \leq 0 \qquad \text{and} \qquad \mathbf{a}^* \cdot \mathbf{v}^{e^{SC}} \leq 1 \tag{11}$$

follow from the fact that this is a Bell inequality for the bound 0 and with algebraic bound 1.

For the opposite inequalities, recall from the previous item that $\mathbf{a}^* \cdot \mathbf{v}^e = \mathsf{CF}(e)$. Thus, we have:

$$\begin{aligned} & \mathbf{a}^* \cdot \mathbf{v}^e = \mathsf{CF}(e) \\ \Leftrightarrow & \qquad \{ \text{ decomposition of eq. (10) and linearity } \} \\ & \mathsf{NCF}(e)\mathbf{a}^* \cdot \mathbf{v}^{e^{NC}} + \mathsf{CF}(e)\mathbf{a}^* \cdot \mathbf{v}^{e^{SC}} = \mathsf{CF}(e) \\ \Leftrightarrow & \\ & \mathsf{NCF}(e)\mathbf{a}^* \cdot \mathbf{v}^{e^{NC}} + \mathsf{CF}(e)(\mathbf{a}^* \cdot \mathbf{v}^{e^{SC}} - 1) = 0 \end{aligned}$$

But the left-hand side now is a convex combination of two numbers which we know to be non-positive from eq. (11). This can only be equal to zero when we have

$$\mathsf{NCF}(e)\mathbf{a}^* \cdot \mathbf{v}^{e^{NC}} = 0 \qquad \text{and} \qquad \mathsf{CF}(e)(\mathbf{a}^* \cdot \mathbf{v}^{e^{SC}} - 1) = 0$$

hence we have that

$$\mathbf{a}^* \cdot \mathbf{v}^{e^{NC}} = 0 \qquad \text{and} \qquad \mathbf{a}^* \cdot \mathbf{v}^{e^{SC}} = 1$$

as long as $\mathsf{NCF}(e) \neq 0$ and $\mathsf{CF}(e) \neq 0$, respectively. □

Remark. Incidentally, note that no use was made of the assumption that $\mathsf{CF}(e) > 0$ in the proof of item (ii), and similarly for the first part of item (iii). However, when $\mathsf{CF}(e) = 0$, i.e. when the model e is non-contextual, it may happen that the Bell inequality obtained by the method described is trivial, in the sense that it is satisfied not only by all non-contextual models but also by all no-signalling ones. In that situation, it is an inequality (in fact, an equality) defining the affine subspace of vectors corresponding to no-signalling models. Indeed, this necessarily happens for models in the relative interior of the non-contextual polytope, for such a model cannot saturate any proper Bell inequality that separates the non-contextual from general no-signalling empirical models. Therefore, even though the statement would strictly remain true given our definitions, we have chosen to include the extra assumption to avoid it being misconstrued.

As noted in the main text, decompositions of the form $e = \mathsf{NCF}(e)e^{NC} + \mathsf{CF}(e)e^{SC}$ are not necessarily unique. This can happen when there is a face of the no-signalling polytope consisting only of strongly contextual models (i.e. whose vertices are all strongly contextual) that is parallel to a face of the non-contextual polytope. If these faces have dimension at least 1 and the model e lies in between them, then any line going through e and intersecting the two faces determines two models e^{NC} and e^{SC}, corresponding to the intersections. For all these lines, the value of λ in the decomposition $e = \lambda e^{NC} + (1-\lambda)e^{SC}$ will be the same. An example of non-uniqueness in the $(3, 2, 2)$ Bell scenario is given by the models in Table II. On the other hand, non-uniqueness cannot arise when the scenario is such that there are no two adjacent strongly contextual vertices of the no-signalling polytope, and hence no face of the polytope consisting solely of strongly contextual models: this is the case, for example, for the $(2, 2, 2)$ Bell scenario.

B. Computational explorations

Computational tools in the form of a *Mathematica* package have been implemented, which can compute:

1. the empirical model arising from any quantum state and any sets of compatible measurements;
2. the incidence matrix for any measurement scenario;
3. the contextual fraction of any empirical model using LP (3);
4. the Bell inequality of Theorem 1, using the dual LP (4) (under change of variables).

We stress that these tools are completely general: they can be applied to any pure or mixed quantum state in any Hilbert space and to any sets of compatible observables in that space, including Bell scenarios as a special case.

Equatorial measurements on the Bell state $|\Phi^+\rangle$.—As an example of how the package can be used, we consider a family of empirical models that can be obtained by considering local measurements on the two-qubit Bell state

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$$

Recall that projective measurements on a qubit can equivalently be represented by a point on the Bloch sphere. Suppose that we allow the same two local measurements on each qubit, and that these are equatorial on the Bloch sphere, parametrised by angles ϕ_1 and ϕ_2 as in Figure 3. We assume these angles are in the interval $[0, \pi)$ since ϕ and $\phi + \pi$ correspond to the same measurement up to relabelling the outcomes. One such model is the Bell–CHSH model from Table I, which is obtained when

$$(\phi_1, \phi_2) = (0, \pi/3) \ .$$

We can use the package to plot the non-contextual fraction of the resulting models as a function of ϕ_1 and ϕ_2

TABLE II: Empirical models e_1^{SC}, e_1^{NC}, e_2^{SC}, and e_2^{NC} in a (3,2,2) Bell scenario, illustrating non-uniqueness of decomposition (2). The models on the left are strongly contextual, those on the right are non-contextual, and ${}^1\!/\!{}_2 e_1^{NC} + {}^1\!/\!{}_2 e_1^{SC} = {}^1\!/\!{}_2 e_2^{NC} + {}^1\!/\!{}_2 e_2^{SC}$ is an empirical model with contextual fraction 1/2.

e_1^{SC}

A	B	C	000	001	010	011	100	101	110	111
a_1	b_1	c_1	0	1/4	1/4	0	1/4	0	0	1/4
a_1	b_1	c_2	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
a_1	b_2	c_1	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
a_1	b_2	c_2	1/4	0	0	1/4	0	1/4	1/4	0
a_2	b_1	c_1	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
a_2	b_1	c_2	1/4	0	0	1/4	0	1/4	1/4	0
a_2	b_2	c_1	1/4	0	0	1/4	0	1/4	1/4	0
a_2	b_2	c_2	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

e_1^{NC}

A	B	C	000	001	010	011	100	101	110	111
a_1	b_1	c_1	1/16	3/16	3/16	1/16	3/16	1/16	1/16	3/16
a_1	b_1	c_2	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
a_1	b_2	c_1	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
a_1	b_2	c_2	3/16	1/16	1/16	3/16	1/16	3/16	3/16	1/16
a_2	b_1	c_1	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8
a_2	b_1	c_2	3/16	1/16	1/16	3/16	1/16	3/16	3/16	1/16
a_2	b_2	c_1	3/16	1/16	1/16	3/16	1/16	3/16	3/16	1/16
a_2	b_2	c_2	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

e_2^{SC}

A	B	C	000	001	010	011	100	101	110	111
a_1	b_1	c_1	0	1/8	1/8	0	3/8	0	0	3/8
a_1	b_1	c_2	0	1/8	1/8	0	1/8	1/4	1/4	1/8
a_1	b_2	c_1	0	1/8	1/8	0	1/8	1/4	1/4	1/8
a_1	b_2	c_2	1/8	0	0	1/8	0	3/8	3/8	0
a_2	b_1	c_1	1/8	0	0	1/8	1/4	1/8	1/8	1/4
a_2	b_1	c_2	1/8	0	0	1/8	0	3/8	3/8	0
a_2	b_2	c_1	1/8	0	0	1/8	0	3/8	3/8	0
a_2	b_2	c_2	0	1/8	1/8	0	1/8	1/4	1/4	1/8

e_2^{NC}

A	B	C	000	001	010	011	100	101	110	111
a_1	b_1	c_1	1/16	5/16	5/16	1/16	1/16	1/16	1/16	1/16
a_1	b_1	c_2	1/4	1/8	1/8	1/4	1/8	0	0	1/8
a_1	b_2	c_1	1/4	1/8	1/8	1/4	1/8	0	0	1/8
a_1	b_2	c_2	5/16	1/16	1/16	5/16	1/16	1/16	1/16	1/16
a_2	b_1	c_1	1/8	1/4	1/4	1/8	0	1/8	1/8	0
a_2	b_1	c_2	5/16	1/16	1/16	5/16	1/16	1/16	1/16	1/16
a_2	b_2	c_1	5/16	1/16	1/16	5/16	1/16	1/16	1/16	1/16
a_2	b_2	c_2	1/4	1/8	1/8	1/4	1/8	0	0	1/8

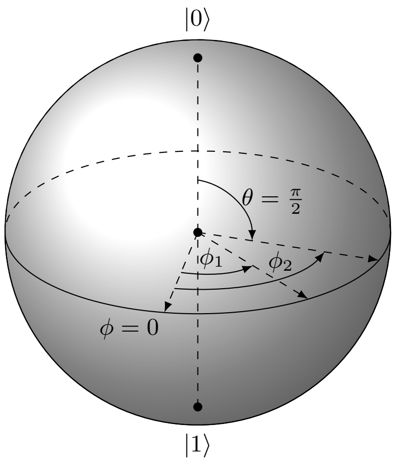


FIG. 3: Equatorial measurements at ϕ_1 and ϕ_2 on the Bloch sphere.

(Figure 1–(a)). It is interesting to note that the Bell–CHSH model from Table I does not achieve the maximum possible degree of contextuality among these models. Instead, the maxima of the plot occur when

$$\{\phi_1, \phi_2\} \in \left\{ \left\{ \frac{\pi}{8}, \frac{5\pi}{8} \right\}, \left\{ \frac{7\pi}{8}, \frac{3\pi}{8} \right\} \right\} .$$

All of the corresponding empirical models take the form of Table III, with

$$p = \frac{\sqrt{2}+2}{8}.$$

TABLE III: Empirical models corresponding to maxima of the plot shown in Figure 1–(a), where $p = {}^{\sqrt{2}+2}\!/\!{}_8$. These achieve the Tsirelson violation of the CHSH inequality.

A	B	00	01	10	11
a_1	b_1	p	$(1/2 - p)$	$(1/2 - p)$	p
a_1	b_2	$(1/2 - p)$	p	p	$(1/2 - p)$
a_2	b_1	$(1/2 - p)$	p	p	$(1/2 - p)$
a_2	b_2	$(1/2 - p)$	p	p	$(1/2 - p)$

These can easily be seen to achieve the Tsirelson violation of the CHSH inequality. Note that none of these models are strongly contextual: this observation provided one motivation to look for proofs that Bell states cannot witness logical forms of contextuality with a finite number of measurements [13, 15], although they do so at the limit where the number of measurement settings tends to infinity [27].

It may seem surprising at first that the degree of contextuality of the empirical models is not constant with respect to the relative angle $(\phi_2 - \phi_1)$ between measurements, a fact that is apparent from the plot. For example, the empirical model obtained when $(\phi_1, \phi_2) = (0, \pi/2)$ is local, but if these values are shifted by $\pi/8$ the resulting empirical model achieves the maximum violation of the CHSH inequality. As it happens, a rotation by Φ around the Z-axis for each of the qubits is described by

$$\begin{pmatrix} e^{-i\Phi/2} & 0 \\ 0 & e^{i\Phi/2} \end{pmatrix} \otimes \begin{pmatrix} e^{-i\Phi/2} & 0 \\ 0 & e^{i\Phi/2} \end{pmatrix}$$

$$= \begin{pmatrix} e^{-i\Phi} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\Phi} \end{pmatrix} . \quad (12)$$

If one equivalently thinks of leaving the measurements fixed and applying the rotations to the state instead, note that this indeed introduces a relative phase of 2Φ between the terms in $|\Phi^+\rangle$, explaining the difference in the resulting empirical models.

Equatorial measurements on n-partite GHZ states.— We can consider similar families of models for the n-partite GHZ states [29], given for each $n > 2$ by:

$$|\psi_{\mathsf{GHZ}(n)}\rangle = \frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}} \quad (13)$$

Note that with $n = 2$ this would simply reduce to the $|\Phi^+\rangle$ Bell state. For $n > 2$, Mermin considered the situation where each each of the n parties can perform Pauli X or Y measurements, and gave logical proofs of strong contextuality (nonlocality) via a parity argument which he called 'all versus nothing' [30]. In [31] it was shown that a very general form of all-*vs*-nothing contextuality implies strong contextuality.

As before, we choose any two equatorial measurements on the Bloch sphere, and make these same two measurements available at each qubit. For $|\psi_{\mathsf{GHZ}(3)}\rangle$ and $|\psi_{\mathsf{GHZ}(4)}\rangle$ we obtain the plots shown in Figures 1–(b) and 1–(c), respectively. The maxima of the plot for the tripartite state reach $\mathsf{CF}(e) = 1$, indicating strong contextuality, and occur when

$$\{\phi_1, \phi_2\} \in \left\{ \left\{\frac{\pi}{2}, 0\right\}, \left\{\frac{2\pi}{3}, \frac{\pi}{6}\right\}, \left\{\frac{5\pi}{6}, \frac{\pi}{3}\right\} \right\} . \quad (14)$$

Of course, $(\phi_1, \phi_2) = (\pi/2, 0)$ correspond to the Pauli measurements Y and X, respectively, yielding the usual GHZ–Mermin model. The empirical models corresponding to other maxima are identical up to re-labelling, and so these provide alternative sets of measurements that can be made on the GHZ state and still lead to the familiar all-*vs*-nothing argument for contextuality.

The situation is similar for $n = 4$, in which maxima of $\mathsf{CF}(e) = 1$ are seen to occur at

$$\{\phi_1, \phi_2\} \in \left\{ \left\{\frac{\pi}{2}, 0\right\}, \left\{\frac{5\pi}{8}, \frac{\pi}{8}\right\}, \left\{\frac{3\pi}{4}, \frac{\pi}{4}\right\}, \left\{\frac{7\pi}{8}, \frac{3\pi}{8}\right\} \right\} . \quad (15)$$

We can see a pattern beginning to emerge in (14) and (15), which leads to the following proposition.

Proposition 4. *Equatorial measurements at*

$$(\phi_1, \phi_2) \in \left\{ \left(\frac{(n+k)\,\pi}{2n}, \frac{k\,\pi}{2n} \right) \mid 0 \le k < n \right\}$$

on each qubit of the $|\psi_{\mathsf{GHZ}(n)}\rangle$ *state give rise (up to re-labelling of measurements and outcomes) to the strongly contextual GHZ–Mermin n-partite empirical model.*

Proof. First, we know that this holds for $k = 0$, since in that case we simply have local Pauli X and Y measurements, the measurements prescribed in Mermin's argument. For $0 < k < n$, we can apply at each qubit a rotation around the Z axis by the phase $\Phi := {}^{k\pi}/_{2n}$, so that we continue to deal with the X and Y measurements. It is necessary, however, to take account of the relative phase introduced by these operations on the n-qubit state. By generalising (12) it is clear that the state obtained from $|\psi_{\mathsf{GHZ}(n)}\rangle$ after rotating each qubit by Φ is

$$|\mathrm{GHZ}(n, \Phi)\rangle = \frac{1}{\sqrt{2}} \left(|0\cdots 0\rangle + e^{in\Phi} |1\cdots 1\rangle \right) .$$

Notice that for the relevant values of Φ the relative phase is a multiple of ${}^{\pi}/_{2}$. Rotating only one of the qubits by ${}^{\pi}/_{2}$ an appropriate number of times brings us back to the state $|\psi_{\mathsf{GHZ}(n)}\rangle$, while each step changes the measurements X and Y to $-Y$ and X, respectively. Since $-Y$ is just the measurement Y with a relabelling of the outcomes, the whole change simply amounts to relabelling of measurements and their outcomes. □

C. Monotonicity (Proof of Theorem 2)

First, we formally define the operations of translation of measurements and coarse-graining of outcomes. For the former, given an empirical model e on the measurement scenario $\langle X', \mathcal{M}', O\rangle$, a second measurement scenario $\langle X, \mathcal{M}, O\rangle$, and a function $f : X \longrightarrow X'$ that preserves contexts, i.e. such that $C \in \mathcal{M}$ implies $f(C) \subseteq C'$ for some $C' \in \mathcal{M}'$, we define the empirical model f^*e in $\langle X, \mathcal{M}, O\rangle$ by pulling e back along the map f: for each $C \in \mathcal{M}$ and $s \in O^C$,

$$(f^*e)_C(s) := \sum_{t \in O^{f(C)}, t\circ f|_C = s} e_{f(C)}(t) .$$

For the latter, given an empirical model e on the measurement scenario $\langle X, \mathcal{M}, O'\rangle$ and a function $h : O' \longrightarrow O$, we define an empirical model e/f on the scenario $\langle X, \mathcal{M}, O\rangle$ as follows: for each $C \in \mathcal{M}$ and $s \in O^C$

$$(e/h)_C(s) := \sum_{s' \in O'^C, h\circ s' = s} e_C(s') .$$

Two of the three combining operations were already introduced formally in the main text. As for mixing, it has the obvious definition: given empirical models e and e' in $\langle X, \mathcal{M}, O\rangle$ and a weight $\lambda \in [0, 1]$, the empirical model $\lambda e + (1 - \lambda)e'$ is defined, for each $C \in \mathcal{M}$ and $s \in O^C$, as

$$(\lambda e + (1-\lambda)e')_C(s) := \lambda\, e_C(s) \; + \; (1-\lambda)\, e'_C(s) .$$

We illustrate the use of the operations to construct local models for Bell scenarios, as mentioned in the main text. We start from the generator $G = \{m \longmapsto *\}$, which sends a single measurement, deterministically, to a single outcome. We define functions $f_i :: * \longmapsto i$, $i \in \{1, \ldots, l\}$ to relabel outcomes of measurements into the set $O = \{1, \ldots, l\}$. A deterministic model for a single agent with k measurements $\{m_1, \ldots, m_k\}$, where measurement m_j is assigned outcome i_j, is described (up to isomorphic relabelling) by $\&_{j=1}^{k} G/f_{i_j}$. Let A be one such model, and B another. Then the corresponding bipartite local model is described by $A \otimes B$. This obviously generalizes to any number of parties. Finally, any local model can be expressed as a mixture of deterministic local models.

We now present a detailed proof of Theorem 2 which states the monotonicity properties of the measure of contextuality CF with respect to these five operations.

Theorem 2. *The contextual fraction is invariant under relabellings, and non-increasing under translation of measurements and coarse-graining of outcomes. For the combining operations, it satisfies the following properties:*

- $\mathsf{CF}(\lambda e + (1-\lambda)e') \leq \lambda \mathsf{CF}(e) + (1-\lambda)\mathsf{CF}(e)$
- $\mathsf{CF}(e \,\&\, e') = \max\{\mathsf{CF}(e), \mathsf{CF}(e')\}$
- $\mathsf{CF}(e \otimes e') = \mathsf{CF}(e) + \mathsf{CF}(e') - \mathsf{CF}(e)\mathsf{CF}(e')$

Proof. We shall prove the equivalent statements in terms of NCF instead of CF.

Translation of measurements. Let e be a model in $\langle X', \mathcal{M}', O\rangle$ and $f : X \longrightarrow X'$ context-preserving. $\mathsf{NCF}(e)$ is the maximal weight of a subprobability distribution b on global assignments $O^{X'}$ such that $b|_{C'} \leq e_{C'}$ for any $C' \in \mathcal{M}'$. Let b_e be such a probability distribution of maximal weight i.e. the corresponding vector $\mathbf{b}^*$ is an optimal solution to the LP (3).

Define $f^* b_e$ a subprobability distribution on O^X by, for any $g \in O^X$,

$$(f^* b_e)(g) := \sum_{g' \in O^{X'}, g' \circ f = g} b_e(g') \ . \tag{16}$$

Note that this $f^* b_e$ has the same weight as b_e since each $g' \in O^{X'}$ contributes to a single $g \in O^X$.

For any $C \in \mathcal{M}$ and $s \in O^C$, we have

$$
\begin{aligned}
& f^* b_e|_C(s) \\
= \quad & \{ \text{ definition of marginalisation } \} \\
& \sum_{g \in O^X, g|_C = s} f^* b_e(g) \\
= \quad & \{ \text{ definition of } f^* b_e \text{, eq. (16) } \} \\
& \sum_{g \in O^X, g|_C = s} \ \sum_{g' \in O^{X'}, g' \circ f = g} b_e(g') \\
= \quad & \\
& \sum_{g' \in O^{X'}, (g' \circ f)|_C = s} b_e(g') \\
= \quad & \\
& \sum_{g' \in O^{X'}, g'|_{f(C)} \circ f|_C = s} b_e(g') \\
= \quad & \\
& \sum_{t \in O^{f(C)}, t \circ f|_C = s} \ \sum_{g' \in O^{X'}, g'|_{f(C)} = t} b_e(g') \\
= \quad & \{ \text{ definition of marginalisation } \} \\
& \sum_{t \in O^{f(C)}, t \circ f|_C = s} b_e|_{f(C)}(t) \\
\leq \quad & \{ \ b_e|_{f(C)} \text{ is subdistribution of } e_{f(C)} \ \} \\
& \sum_{t \in O^{f(C)}, t \circ f|_C = s} e_{f(C)}(t) \\
= \quad & \{ \text{ definition of } f^* e \ \} \\
& (f^* e)_C(s)
\end{aligned}
$$

i.e. $f^* b_e$ corresponds to a feasible solution to the LP for model $f^* e$. Hence,

$$\mathsf{NCF}(f^* e) \geq w(f^* b_e) = w(b_e) = \mathsf{NCF}(e) \ .$$

Coarse-graining of outcomes. Let e be a model in $\langle X, \mathcal{M}, O'\rangle$ and $h : O' \longrightarrow O$. Again, write b_e for a subprobability distribution on O'^X of maximal weight $\mathsf{NCF}(e)$ satisfying $b_e|_C \leq e_C$ for all $C \in \mathcal{M}$.

Define b_e/h a subprobability distribution on O^X by, for any $g \in O^X$,

$$(b_e/h)(g) := \sum_{g' \in O'^X, h \circ g' = g} b_e(g') \ . \tag{17}$$

Similarly, note that b_e/h has the same weight as b_e since each $g' \in O'^X$ contributes to a single $g \in O^X$.

For any $C \in \mathcal{M}$ and $s \in O^C$, we have

$$
\begin{aligned}
& (b_e/h)|_C(s) \\
= \quad & \{ \text{ definition of marginalisation } \} \\
& \sum_{g \in O^X, g|_C = s} (b_e/h)(g) \\
= \quad & \{ \text{ definition of } b_e/h \text{, eq. (17) } \} \\
& \sum_{g \in O^X, g|_C = s} \ \sum_{g' \in O'^X, h \circ g' = g} b_e(g') \\
= \quad & \\
& \sum_{g' \in O'^X, (h \circ g')|_C = s} b_e(g') \\
= \quad & \\
& \sum_{g' \in O'^X, h \circ g'|_C = s} b_e(g') \\
= \quad & \\
& \sum_{s' \in O'^C, h \circ s' = s} \ \sum_{g' \in O'^X, g'|_C = s'} b_e(g')
\end{aligned}
$$

$$
\begin{aligned}
&= \quad \{ \text{ definition of marginalisation } \} \\
&\quad \sum_{s' \in O'^C, h \circ s' = s} b_e|_C(s') \\
&\leq \quad \{ \; b_e|_C \text{ is subdistribution of } e_C \; \} \\
&\quad \sum_{s' \in O'^C, h \circ s' = s} e_C(s') \\
&= \quad \{ \text{ definition of } e/h \} \\
&\quad (e/h)_C(s)
\end{aligned}
$$

i.e. b_e/h corresponds to a feasible solution to the LP for model e/h. Therefore,

$$\mathsf{NCF}(e/h) \geq w(b_e/h) = w(b_e) = \mathsf{NCF}(e) \ .$$

Mixing. Let e_1 and e_2 be models in $\langle X, \mathcal{M}, O\rangle$ and $\lambda \in [0,1]$. Note that in terms of vector representation we have $\mathbf{v}^{\lambda e_1 + (1-\lambda)e_2} = \lambda \mathbf{v}^{e_1} + (1-\lambda)\mathbf{v}^{e_2}$.

Let $\mathbf{b}_i^*$ be an optimal solution to the LP (3) relative to model e_i ($i \in \{1,2\}$), and set $\mathbf{b} := \lambda \mathbf{b}_1^* + (1-\lambda)\mathbf{b}_2^*$. Then, $\mathbf{b} \geq 0$ follows from non-negativity of $\mathbf{b}_1^*$ and $\mathbf{b}_2^*$, and moreover

$$
\begin{aligned}
&\quad \mathbf{M}\,\mathbf{b} \\
&= \\
&\quad \mathbf{M}\left(\lambda \mathbf{b}_1^* + (1-\lambda)\mathbf{b}_2^*\right) \\
&= \quad \{ \text{ linearity } \} \\
&\quad \lambda \mathbf{M}\,\mathbf{b}_1^* + (1-\lambda)\mathbf{M}\,\mathbf{b}_2^* \\
&\leq \quad \{ \text{ feasibility of } \mathbf{b}_i^*,\ \mathbf{M}\,\mathbf{b}_i^* \leq \mathbf{v}^{e_i} \} \\
&\quad \lambda \mathbf{v}^{e_1} + (1-\lambda)\mathbf{v}^{e_2} \\
&= \\
&\quad \mathbf{v}^{\lambda e_1 + (1-\lambda)e_2} \ .
\end{aligned}
$$

This means that $\mathbf{b}$ is a feasible solution to the LP relative to the model $\lambda e_1 + (1-\lambda)e_2$, achieving the following value of the objective function:

$$
\begin{aligned}
&\quad \mathbf{1} \cdot \mathbf{b} \\
&= \\
&\quad \mathbf{1} \cdot (\lambda \mathbf{b}_1^* + (1-\lambda)\mathbf{b}_2^*) \\
&= \quad \{ \text{ linearity } \} \\
&\quad \lambda(\mathbf{1} \cdot \mathbf{b}_1^*) + (1-\lambda)(\mathbf{1} \cdot \mathbf{b}_2^*) \\
&= \quad \{ \text{ optimality of } \mathbf{b}_i^*,\ \mathbf{1} \cdot \mathbf{b}_i^* = \mathsf{NCF}(e_i) \} \\
&\quad \lambda\,\mathsf{NCF}(e_1) + (1-\lambda)\,\mathsf{NCF}(e_2) \ .
\end{aligned}
$$

Since $\mathsf{NCF}(\lambda e_1 + (1-\lambda)e_2)$ is the optimal (maximum) value for this primal LP, we have that

$$\mathsf{NCF}(\lambda e_1 + (1-\lambda)e_2) \ \geq\ \lambda\,\mathsf{NCF}(e_1) \ +\ (1-\lambda)\,\mathsf{NCF}(e_2)$$

as desired.

Product. Let e_1 and e_2 be models on $\langle X_1, \mathcal{M}_1, O\rangle$ and $\langle X_2, \mathcal{M}_2, O\rangle$, respectively. Note that global assignments for the scenario $\langle X_1 \sqcup X_2, \mathcal{M}_1 \star \mathcal{M}_2, O\rangle$ are in bijective correspondence with tuples $\langle g_1, g_2\rangle$ where $g_i : X_i \longrightarrow O$ is a global assignment for $\langle X_i, \mathcal{M}_i, O\rangle$ ($i \in \{1,2\}$). Similarly, contexts $C \in \mathcal{M}_1 \star \mathcal{M}_2$ are those of the form $C = C_1 \sqcup C_2$ with $C_i \in \mathcal{M}_i$, hence they are in bijective correspondence with pairs $\langle C_1, C_2\rangle$ of contexts with $C_i \in \mathcal{M}_i$. Consequently, local assignments $\langle C_1 \sqcup C_2 \in \mathcal{M}_1 \star \mathcal{M}_2, s : C_1 \sqcup C_2 \longrightarrow O\rangle$ of the scenario $\langle X_1 \sqcup X_2, \mathcal{M}_1 \star \mathcal{M}_2, O\rangle$ are in bijective correspondence with pairs of local assignments for each of the scenarios, i.e. pairs $\langle\langle C_1, s_1\rangle, \langle C_2, s_2\rangle\rangle$ with $\langle C_i \in \mathcal{M}_i, s_i : C_i \longrightarrow O\rangle$. We use these equivalent representations to index the representation of empirical models as vectors, the incidence matrix, etc. for the scenario $\langle X_1 \sqcup X_2, \mathcal{M}_1 \star \mathcal{M}_2, O\rangle$.

Observe that, in terms of the vector representations, the product empirical model $e_1 \otimes e_2$ is concisely written as $\mathbf{v}^{e_1 \otimes e_2} = \mathbf{v}^{e_1} \otimes \mathbf{v}^{e_2}$, since for any local assignments $\langle C_i \in \mathcal{M}_i, s_i : C_i \longrightarrow O\rangle$,

$$\mathbf{v}^{e_1 \otimes e_2}[\langle\langle C_1, s_1\rangle, \langle C_2, s_2\rangle\rangle] = \mathbf{v}^{e_1}[\langle C_1, s_1\rangle]\mathbf{v}^{e_2}[\langle C_2, s_2\rangle] \ .$$

Moreover, if $\mathbf{M}_1$ and $\mathbf{M}_2$ are the incidence matrices for each of the measurement scenarios, then the incidence matrix $\mathbf{M}$ for the scenario $\langle X_1 \sqcup X_2, \mathcal{M}_1 \star \mathcal{M}_2, O\rangle$ is precisely $\mathbf{M} = \mathbf{M}_1 \otimes \mathbf{M}_2$ since, for global assignments $g_i : X_i \longrightarrow O$ and local assignments $\langle C_i \in \mathcal{M}_i, s_i : C_i \longrightarrow O\rangle$,

$$
\begin{aligned}
&\mathbf{M}[\langle\langle C_1, s_1\rangle, \langle C_2, s_2\rangle\rangle, \langle g_1, g_2\rangle] \\
&= \begin{cases} 1 & \text{if } \langle g_1, g_2\rangle|_{C_1 \sqcup C_2} = \langle s_1, s_2\rangle; \\ 0 & \text{otherwise.} \end{cases} \\
&= \begin{cases} 1 & \text{if } g_1|_{C_1} = s_1 \text{ and } g_2|_{C_2} = s_2; \\ 0 & \text{otherwise.} \end{cases} \\
&= \mathbf{M}_1[\langle C_1, s_1\rangle, g_1]\,\mathbf{M}_2[\langle C_2, s_2\rangle, g_2] \ .
\end{aligned}
$$

Let $\mathbf{b}_1^*$ and $\mathbf{b}_2^*$ be optimal solutions to the primal LP (3) relative to e_1 and e_2, respectively, and set $\mathbf{b} := \mathbf{b}_1^* \otimes \mathbf{b}_2^*$. Then we have $\mathbf{b} \geq \mathbf{0}$ from the non-negativity of the $\mathbf{b}_i^*$, and moreover

$$
\begin{aligned}
&\quad \mathbf{M}\,\mathbf{b} \\
&= \\
&\quad (\mathbf{M}_1 \otimes \mathbf{M}_2)\,(\mathbf{b}_1^* \otimes \mathbf{b}_2^*) \\
&= \\
&\quad (\mathbf{M}_1\,\mathbf{b}_1^*) \otimes (\mathbf{M}_2\,\mathbf{b}_2^*) \\
&\leq \quad \{ \text{ by feasibility of } \mathbf{b}_i^*,\ \mathbf{M}_i\,\mathbf{b}_i^* \leq \mathbf{v}^{e_i} \} \\
&\quad \mathbf{v}^{e_1} \otimes \mathbf{v}^{e_2} \\
&= \\
&\quad \mathbf{v}^{e_1 \otimes e_2} \ ,
\end{aligned}
$$

hence $\mathbf{b}$ is a feasible solution to the primal LP relative to $e_1 \otimes e_2$, achieving the following value of the objective function:

$$
\begin{aligned}
& \mathbf{1} \cdot \mathbf{b} \\
= \quad & \\
& (\mathbf{1} \otimes \mathbf{1}) \cdot (\mathbf{b}_1^* \otimes \mathbf{b}_2^*) \\
= \quad & \\
& (\mathbf{1} \cdot \mathbf{b}_1^*)(\mathbf{1} \cdot \mathbf{b}_2^*) \\
= \quad & \quad \{ \text{ by optimality of } \mathbf{b}_i^*,\ \mathbf{1} \cdot \mathbf{b}_i^* = \mathsf{NCF}(e_1) \ \} \\
& \mathsf{NCF}(e_1)\mathsf{NCF}(e_2) \ .
\end{aligned}
$$

Since $\mathsf{NCF}(e_1 \otimes e_2)$ is the optimal (maximum) value for this primal LP, we have that

$$\mathsf{NCF}(e_1 \otimes e_2) \geq \mathsf{NCF}(e_1)\mathsf{NCF}(e_2) \ .$$

For the opposite inequality, we follow an analogous argument using the dual LP (4). Let $\mathbf{y}_1^*$ and $\mathbf{y}_2^*$ be optimal solutions to the dual LP for e_1 and e_2, respectively, and set $\mathbf{y} := \mathbf{y}_1^* \otimes \mathbf{y}_2^*$. Then we have $\mathbf{y} \geq 0$ from the non-negativity of the $\mathbf{y}_i^*$, and moreover

$$
\begin{aligned}
& \mathbf{M}^\intercal \mathbf{y} \\
= \quad & \\
& (\mathbf{M}_1^\intercal \otimes \mathbf{M}_2^\intercal)\, (\mathbf{y}_1^* \otimes \mathbf{y}_2^*) \\
= \quad & \\
& (\mathbf{M}_1^\intercal \mathbf{y}_1^*) \otimes (\mathbf{M}_2^\intercal \mathbf{y}_2^*) \\
\geq \quad & \quad \{ \text{ by feasibility of } \mathbf{y}_i^*,\ \mathbf{M}_i^\intercal \mathbf{y}_i^* \geq \mathbf{1} \ \} \\
& \mathbf{1} \otimes \mathbf{1} \\
= \quad & \\
& \mathbf{1}
\end{aligned}
$$

hence $\mathbf{y}$ is a feasible solution to the dual LP relative to $e_1 \otimes e_2$, achieving the following value of the objective function:

$$
\begin{aligned}
& \mathbf{y} \cdot \mathbf{v}^{e_1 \otimes e_2} \\
= \quad & \\
& (\mathbf{y}_1^* \otimes \mathbf{y}_2^*) \cdot (\mathbf{v}^{e_1} \otimes \mathbf{v}^{e_2}) \\
= \quad & \\
& (\mathbf{y}_1^* \cdot \mathbf{v}^{e_1})(\mathbf{y}_2^* \cdot \mathbf{v}^{e_2}) \\
= \quad & \quad \{ \text{ by optimality of } \mathbf{y}_i^*,\ \mathbf{y}_i^* \cdot \mathbf{v}^{e_i} = \mathsf{NCF}(e_i) \ \} \\
& \mathsf{NCF}(e_1)\mathsf{NCF}(e_2) \ .
\end{aligned}
$$

Since $\mathsf{NCF}(e_1 \otimes e_2)$ is the optimal (minimum) value for this dual LP, we have that

$$\mathsf{NCF}(e_1 \otimes e_2) \leq \mathsf{NCF}(e_1)\mathsf{NCF}(e_2) \ .$$

Choice. Let e_1 and e_2 be models on $\langle X_1, \mathcal{M}_1, O\rangle$ and $\langle X_2, \mathcal{M}_2, O\rangle$, respectively, and consider the model $e_1 \,\&\, e_2$ on $\langle X_1 \sqcup X_2, \mathcal{M}_1 \sqcup \mathcal{M}_2, O\rangle$. Write $b_{e_1\&e_2}$ for a subprobability distribution on global assignments $O^{X_1 \sqcup X_2} \cong O^{X_1} \times O^{X_2}$ of maximal weight $\mathsf{NCF}(e_1 \,\&\, e_2)$ that satisfies $b_{e_1\&e_2}|_C \leq (e_1 \,\&\, e_2)_C$ for any $C \in \mathcal{M}_1 \sqcup \mathcal{M}_2$.

For $i \in \{1,2\}$, define a subprobability distribution b_i on O^{X_i} by $b_i := b_{e_1\&e_2}|_{X_i}$. Note that each b_i has the same weight as $b_{e_1\&e_2}$. Then, for any $C \in \mathcal{M}_i$ and $s \in O^C$, we have

$$
\begin{aligned}
& b_i|_C(s) \\
= \quad & \quad \{ \text{ definition of } b_i \ \} \\
& (b_{e_1\&e_2}|_{X_i})|_C(s) \\
= \quad & \\
& b_{e_1\&e_2}|_C(s) \\
\leq \quad & \quad \{ \text{ by feasibility of } b_{e_1\&e_2} \text{ for } e_1 \,\&\, e_2 \ \} \\
& (e_1 \,\&\, e_2)_C(s) \\
= \quad & \quad \{ \text{ definition of } e_1 \,\&\, e_2 \ \} \\
& (e_i)_C(s)
\end{aligned}
$$

That is, b_i corrresponds to a feasible solution of the LP for the model e_i, implying that

$$\mathsf{NCF}(e_i) \geq w(b_i) = w(b_{e_1\&e_2}) = \mathsf{NCF}(e_1 \,\&\, e_2) \ .$$

Therefore, $\mathsf{NCF}(e_1 \,\&\, e_2) \leq \min\{\mathsf{NCF}(e_1), \mathsf{NCF}(e_2)\}$.

For the opposite inequality, let b_{e_1} and b_{e_2} be subprobability distributions on O^{X_1} and O^{X_2}, respectively, corresponding to optimal solutions to the LP for e_1 and e_2, respectively. The goal is to define a subprobability distribution b on $O^{X_1 \sqcup X_2} \cong O^{X_1} \times O^{X_2}$ with weight $w(b) = \min\{w(b_{e_1}), w(b_{e_2})\}$ such that

$$b|_{X_1} \leq b_{e_1} \qquad \text{and} \qquad b|_{X_2} \leq b_{e_2} \ . \tag{18}$$

This condition guarantees that this is a feasible solution to the LP for $e_1 \& e_2$, since, for any context $C \in \mathcal{M}_1 \sqcup \mathcal{M}_2$, writing i for the component to which C belongs, we have:

$$
\begin{aligned}
& b|_C \\
= \quad & \\
& b|_{X_i}|_C \\
\leq \quad & \quad \{ \text{ by eq. (18): } b|_{X_i} \leq b_{e_i} \ \} \\
& b_{e_i}|_C \\
\leq \quad & \quad \{ \text{ by feasibility of } b_{e_i} \text{ for } e_i \ \} \\
& (e_i)_C \\
= \quad & \quad \{ \text{ definition of } e_1 \,\&\, e_2 \ \} \\
& (e_1 \,\&\, e_2)_C
\end{aligned}
$$

This in turn implies that

$$\mathsf{NCF}(e_1 \,\&\, e_2) \geq w(b)$$

$$= \min\{w(b_{e_1}), w(b_{e_2})\}$$
$$= \min\{\mathsf{NCF}(e_1), \mathsf{NCF}(e_2)\} \ .$$

It thus remains to show that this b can be contructed. This is achieved by Lemma 6 proved below. □

Lemma 5. *Let b_S and b_T be subprobability distributions on sets S and T, respectively, with the same weight w. Then there exists a subprobability distribution b on $S \times T$ with weight w whose marginals are the original subdistributions.*

Proof. Let $S_+ \subseteq S$ and $T_+ \subseteq T$ stand for the (finite) supports of the distributions b_S and b_T, respectively. Choose (any) total orderings of the sets S_+ and T_+, i.e. :

$$S_+ = \{s_1, \dots, s_n\} \qquad \text{and} \qquad T_+ = \{t_1, \dots, t_m\} \ .$$

Define the functions $L_S, R_S : S_+ \longrightarrow [0, w]$ as follows: for each $s_i \in S_+$,

$$L_S(s_i) := \sum_{1 \le j < i} b_S(s_j)$$
$$R_S(s_i) := \sum_{1 \le j \le i} b_S(s_j) \ .$$

Note that $b_S(s) = R_S(s) - L_S(s)$, and moreover that $R_S(s_i) = L_S(s_{i+1})$ for all $i \in \{1, \dots, n-1\}$. R_T and L_T are defined analogously based on the subprobability distribution b_T.

The subprobability distribution b on $S \times T$ is given, for any $s \in S_+$ and $t \in T_+$, as

$$b(s,t) := \min\{R_S(s), R_T(t)\} \ominus \max\{L_S(s), L_T(t)\}$$

where $r \ominus l := \max\{r - l, 0\}$, and by $b(s,t) := 0$ whenever $s \in S \setminus S_+$ or $t \in T \setminus T_+$.

We show that $b|_S = b_S$; the proof that $b|_T = b_T$ is analogous. Note that this also means that $w(b) = w(b_S) = w$.

Each $s \in S_+$ determines a partition of T_+ into the following disjoint subsets (the letters stand for 'to the left', 'overlapping', and 'to the right'):

$$\mathcal{L}_s := \{t \in T_+ \mid R_T(t) \le L_S(s)\}$$
$$\mathcal{O}_s := \{t \in T_+ \mid L_S(s) < R_T(t) \wedge L_T(t) < R_S(s)\}$$
$$\mathcal{R}_s := \{t \in T_+ \mid L_T(t) \ge R_S(s)\}$$

These sets satisfy the property that $l < o < r$ for any $t_l \in \mathcal{L}_s$, $t_o \in \mathcal{O}_s$, and $t_r \in \mathcal{R}_s$. Observe that $b(s,t) \neq 0$ if and only if $t \in \mathcal{O}_s$, and in that case, $b(s,t) = \min\{R_S(s), R_T(t)\} - \max\{L_S(s), L_T(t)\}$.

Since $R_S(s) - L_S(s) = b_S(s) > 0$, one must have $\mathcal{O}_s \neq \emptyset$. So, let t_p and t_u be, respectively, the first and last elements of $\mathcal{O}_s$ (note that we are not excluding the possibility that $\mathcal{O}_s$ has a single element; in that case we merely have $p = u$). Observe that

$$\max\{L_S(s), L_T(t_p)\} = L_S(s) \ ,$$

for otherwise $R_T(t_{p-1}) = L_T(t_p) > L_S(s)$ meaning that $t_{p-1} \notin \mathcal{L}_s$, which would contradict the minimality of t_p in $\mathcal{O}_s$. On the other hand, for any other element of $\mathcal{O}_s$, i.e. for any t_j with $p < j \le u$,

$$\max\{L_S(s), L_T(t_j)\} = L_T(t_j) \ ,$$

for otherwise $L_S(s) > L_T(t_j) \ge R_T(t_p)$ and we would have $t_p \notin \mathcal{O}_s$, a contradiction.

Dually, we have

$$\min\{R_S(s), R_T(t_u)\} = R_S(s) \ ,$$

and

$$\min\{R_S(s), R_T(t_j)\} = R_T(t_j)$$

for any t_j with $p \le j < u$.

Therefore,

$$\begin{aligned}
& b|_S(s) \\
= \ & \sum_{t \in \mathcal{O}_s} b(s,t) \\
= \ & \sum_{p \le j \le u} (\min\{R_S(s), R_T(t_j)\} - \max\{L_S(s), L_T(t_j)\}) \\
= \ & \min\{R_S(s), R_T(t_u)\} + \sum_{p \le j < u} \min\{R_S(s), R_T(t_j)\} \\
& - \sum_{p < j \le u} \max\{L_S(s), L_T(t_j)\} - \max\{L_S(s), L_T(t_p)\} \\
= \ & R_S(s) + \sum_{p \le j < u} R_T(t_j) - \sum_{p < j \le u} L_T(t_j) - L_S(s) \\
= \ & R_S(s) + \sum_{p \le j < u} L_T(t_{j+1}) - \sum_{p < j \le u} L_T(t_j) - L_S(s) \\
= \ & R_S(s) + \sum_{p < j \le u} L_T(t_j) - \sum_{p < j \le u} L_T(t_j) - L_S(s) \\
= \ & R_S(s) - L_S(S) \\
= \ & b_S(s)
\end{aligned}$$

□

Lemma 6. *Let b_S and b_T be subprobability distribution on sets S and T, respectively. Then there exists a subprobability distribution b on $S \times T$ with weight $w(b) = \min\{w(b_S), w(b_T)\}$ such that $b|_S \le b_S$ and $b|_T \le b_T$.*

Proof. If one of the distributions has zero weight, the result is obvious, so let $w(b_S), w(b_T) > 0$. Without loss of generality, assume $w(b_S) \le w(b_T)$ and renormalise b_T

by the shrinking factor $\frac{w(b_S)}{w(b_T)}$, yielding a subprobability distribution b'_T that is a subdistribution of b_T and has the same weight as b_S. The result then follows from Lemma 5. □

D. Contextual fraction and $l2$-MBQC (Proof of Theorem 3)

We recall (and rephrase) from [2] the definitions of measurement-based quantum computation with $\mathbb{Z}_2$-linear classical processing ($l2$-MBQC). A computation of this kind is performed by a parity computer, acting as the classical control to choose measurement settings and processing outcomes, with access to a resource in the form of an empirical model on a multipartite scenario.

Note that in an $(n, 2, 2)$ Bell scenario—i.e. a scenario where n parties can each choose between performing one of two different measurements, which may each yield one of two possible outcomes—the measurement contexts, which comprise a choice of measurement setting for each party, can be represented by a vector in 2^n. Similarly, a joint outcome may also be represented as a vector in 2^n. Therefore, such an empirical model e determines a function $2^n \longrightarrow \mathcal{D}(2^n)$ that associates to each (measurement) vector $\mathbf{q} \in 2^n$ the probability distribution $e_{\mathbf{q}}$ on outcome vectors in 2^n.

An $l2$-MBQC with m bits of input and l bits of output using an n-partite resource consists of:

- a pre-processing $n \times m$ $\mathbb{Z}_2$-matrix $\mathbf{Q}$;
- a post-processing $l \times n$ $\mathbb{Z}_2$-matrix $\mathbf{Z}$;
- an $n \times n$ strictly lower triangular $\mathbb{Z}_2$-matrix $\mathbf{T}$ representing the flow;
- an empirical model on the $(n, 2, 2)$ Bell scenario.

We shall often denote an $l2$-MBQC as a pair $\langle K, e\rangle$ where K is the description of the classical processing (the triple of matrices $\langle \mathbf{Q}, \mathbf{T}, \mathbf{Z}\rangle$) and e the empirical model used as a resource.

Each execution starts with a vector $\mathbf{i} \in 2^m$ of inputs and calculates a vector $\mathbf{o} \in 2^l$ of outputs, using two intermediate vectors $\mathbf{q}, \mathbf{s} \in 2^n$, proceeding according to:

$$\mathbf{q} := \mathbf{Q}\mathbf{i} + \mathbf{T}\mathbf{s} \quad , \quad \mathbf{o} := \mathbf{Z}\mathbf{s} \ , \tag{19}$$

and with $\mathbf{s}$ obtained from $\mathbf{q}$ by sampling the distribution $e_{\mathbf{q}}$ – i.e. by performing the measurements, using the resource. Note that in actual fact the distribution is not sampled *at once*, since parties may be asked to perform their measurements at different times. Indeed, the decision of which measurement to perform on a party may depend on the outcome of the measurement already performed by other parties, as is clear from the way the vector $\mathbf{q}$ is obtained. The reason for $\mathbf{T}$ being strictly lower triangular is exactly to ensure that there is an order in which the computation may be performed (the point is that the j-th component of the vector $\mathbf{q}$, $\mathbf{q}[j]$, i.e. the measurement to be performed by the j-th party, can only depend on the values of $\mathbf{s}[i]$ with $i < j$, i.e. on the outcomes of the measurements performed at parties with index strictly smaller than j). But the fact that e satisfies the no-signalling condition guarantees that a subset of the measurements of a context may be performed unambiguously without the knowledge of what the full context will be.

Note that this execution is probabilistic. As such, the $l2$-MBQC $\langle K, e\rangle$ determines a map $[\![K, e]\!] : 2^m \longrightarrow \mathcal{D}(2^l)$ associating to each input bit string a distribution on output bit strings: given $\mathbf{i} \in 2^m$ and $\mathbf{o} \in 2^l$, $[\![K, e]\!](\mathbf{i})(\mathbf{o})$ is the probability of obtaining output bit string $\mathbf{o}$ when performing the computation $\langle K, e\rangle$ with input bit string $\mathbf{i}$.

The first ingredient necessary to state the result is a measure of success of the computation in implementing a particular objective function. Consider a function $f : 2^m \longrightarrow 2^l$, which one aims to implement. The $l2$-MBQC $\langle K, e\rangle$ evaluates f with worst-case success probability

$$p_S^{\langle K,e\rangle,f} := \min_{\mathbf{i}\in 2^m} [\![K, e]\!](\mathbf{i})(f\mathbf{i}) \ ,$$

and with average success probability

$$\bar{p}_S^{\langle K,e\rangle,f} := \frac{1}{2^m} \sum_{\mathbf{i}\in 2^m} [\![K, e]\!](\mathbf{i})(f\mathbf{i}) \ . \tag{20}$$

We shall omit the upper indices if they are clear from the context. Clearly, we always have $p_S \leq \bar{p}_S$.

We now introduce a measure of the *hardness* of the problem one aims to implement. Since linear functions are the *free* computations in this model, this expresses how much the objective function deviates from being linear. The **average distance** between two Boolean functions $f, g : 2^m \longrightarrow 2^l$ is given by

$$\tilde{d}(f, g) := \frac{1}{2^m} |\{i \in 2^m \mid f(\mathbf{i}) \neq g(\mathbf{i})\}| \ . \tag{21}$$

The average distance of f to the closest linear function is denoted by

$$\tilde{\nu}(f) := \min \left\{ \tilde{d}(f, h) \mid h : 2^m \longrightarrow 2^l \ \mathbb{Z}_2\text{-linear} \right\} \ . \tag{22}$$

We can now present a proof of the result that relates the hardness of the problem, the probability of success, and the contextual fraction of the resource. This sharpens ideas implicit in the proof of [2, Lemma 1].

Theorem 3. *Let $f : 2^m \longrightarrow 2^l$ be a Boolean function and consider an $l2$-MBQC that uses the empirical model e to compute f with average success probability $\bar{p}_S$ over all 2^m possible inputs, and corresponding average failure probability $\bar{p}_F = 1 - \bar{p}_S$. Then, $\bar{p}_F \geq \mathsf{NCF}(e)\tilde{\nu}(f)$.*

Proof. Write $e = \mathsf{NCF}(e)e^{NC} + \mathsf{CF}(e)e^{SC}$, and consider the $l2$-MBQCs $\langle K, e^{NC}\rangle$ and $\langle K, e^{SC}\rangle$ that correspond to using the same classical processing as our $l2$-MBQC of interest but using the empirical model e^{NC} (resp. e^{SC}) instead of e. Then, we have that, for any input $\mathbf{i} \in 2^m$,

$$[\![K, e]\!](\mathbf{i}) = \mathsf{NCF}(e)[\![K, e^{NC}]\!](\mathbf{i}) + \mathsf{CF}(e)[\![K, e^{SC}]\!](\mathbf{i}) \ . \tag{23}$$

and so

$$\begin{aligned}
& \bar{p}_S^{\langle K,e\rangle,f} \\
= \quad & \{ \text{ definition of } \bar{p}_S, \text{ eq. (20) } \} \\
& \frac{1}{2^m} \sum_{\mathbf{i} \in 2^m} [\![K, e]\!](\mathbf{i})(f\mathbf{i}) \\
= \quad & \{ \text{ decomposition of the computation } \langle K, e\rangle, \text{ eq. (23) } \} \\
& \frac{1}{2^m} \sum_{\mathbf{i} \in 2^m} (\mathsf{NCF}(e)[\![K, e^{NC}]\!](\mathbf{i})(f\mathbf{i}) \\
& \qquad + \mathsf{CF}(e)[\![K, e^{SC}]\!](\mathbf{i})(f\mathbf{i})) \\
= \quad & \{ \text{ distributivity and definition of } \bar{p}_S, \text{ eq. (20) (twice) } \} \\
& \mathsf{NCF}(e)\, \bar{p}_S^{\langle K,e^{NC}\rangle,f} \ + \ \mathsf{CF}(e)\, \bar{p}_S^{\langle K,e^{SC}\rangle,f}
\end{aligned}$$

We can bound the probability of success by ignoring what happens when the strongly contextual part of the resource is used (i.e. assuming that it always suceeds in that case):

$$\bar{p}_S^{\langle K,e\rangle,f} \leq \mathsf{NCF}(e)\, \bar{p}_S^{\langle K,e^{NC}\rangle,f} \ + \ \mathsf{CF}(e) \ .$$

Hence,

$$\bar{p}_F^{\langle K,e\rangle,f} \quad \geq \quad \mathsf{NCF}(e)\, \bar{p}_F^{\langle K,e^{NC}\rangle,f} \ .$$

That is, the computation will fail to compute f at least when the non-contextual part of the resource is used and fails to compute f, thus the overall average probability of failure will be at least the average probability of failure of the non-contextual part of the resource multiplied by the weight of this part, i.e. the non-contextual fraction.

It remains to show that the average probability of failure of a non-contextual empirical model is at least $\nu(f)$. One can similarly break down e^{NC} into a convex combination of deterministic non-contextual empirical models,

$$e^{NC} = \sum_{g \in O^X} d(g)\delta_g$$

with d a distribution on O^X, where we write δ_g for the deterministic model determined by the global assignment $g \in O^X$, for which $(\delta_g)_C$ is the delta distribution at $g|_C$. By a similar derivation as above, we have

$$[\![K, e^{NC}]\!](\mathbf{i}) = \sum_{g \in O^X} d(g)\, [\![K, \delta_g]\!](\mathbf{i}) \ ,$$

and then

$$\bar{p}_F^{\langle K,e\rangle,f} = \sum_{g \in O^X} d(g)\, \bar{p}_F^{\langle K,\delta_g\rangle,f} \ .$$

So, it is enough to show that the average probability of failure for any deterministic non-contextual model δ_g is at least $\nu(f)$.

Note that computation by $\langle K, \delta_g\rangle$ is always deterministic, hence there is a function $h^{K,g} : 2^m \longrightarrow 2^l$ such that

$$[\![K, \delta_g]\!](\mathbf{i})(\mathbf{o}) = \delta(h^{K,g}(\mathbf{i}), \mathbf{o}) = \begin{cases} 1 & \text{if } h^{K,g}(\mathbf{i}) = \mathbf{o} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

Moreover, this function $h^{K,g}$ is linear, as shown by Raussendorf [2, Theorem 2]. The proof of this fact essentially comes down to the fact that this function is obtained from composition of several linear functions. The crux of the matter is the transformation from $\mathbf{q}$ to $\mathbf{s}$ (a map $2^n \longrightarrow 2^n$), when determined by a global assignment g as is the case here, can be seen to be built from n maps of type $2 \longrightarrow 2$, each of which is necessarily $\mathbb{Z}_2$-linear, as are all functions of this type. Consequently, one can write $\mathbf{s} = \mathsf{diag}(\mathbf{d})\mathbf{q} + \mathbf{c}$ for some n-component vectors $\mathbf{d}$ and $\mathbf{c}$. Combining this with eq. (19), one obtains

$$\mathbf{o} = \mathbf{Z}\mathbf{s} = \mathbf{Z}(\mathbf{I}_n - \mathsf{diag}(\mathbf{d})\mathbf{T})^{-1}(\mathsf{diag}(\mathbf{d})\mathbf{Q}\mathbf{i} + \mathbf{c}) \ ,$$

where invertibility of $\mathbf{I}_n - \mathsf{diag}(\mathbf{d})\mathbf{T}$ follows from $\mathbf{T}$ being strictly lower triangular. This shows that $\mathbf{o}$ is a given as a linear function of $\mathbf{i}$ when the map from $\mathbf{q}$ to $\mathbf{s}$ is determined by a global assignment; that is, $h^{K,g}$ is linear.

As a consequence, we have that

$$\begin{aligned}
& \bar{p}_F^{\langle K,\delta_g\rangle,f} \\
= \quad & \\
& 1 - \bar{p}_S^{\langle K,\delta_g\rangle,f} \\
= \quad & \{ \text{ definition of } \bar{p}_S, \text{ eq. (20) } \} \\
& 1 - \frac{1}{2^m} \sum_{\mathbf{i} \in 2^m} [\![K, \delta_g]\!](\mathbf{i})(f\mathbf{i}) \\
= \quad & \{ \text{ expanding } [\![K, \delta_g]\!] \text{ by eq. (24) } \} \\
& 1 - \frac{1}{2^m} \sum_{\mathbf{i} \in 2^m} \delta(h^{K,g}(\mathbf{i}), f\mathbf{i}) \\
= \quad & \\
& 1 - \frac{1}{2^m} \left|\{\mathbf{i} \in 2^m \mid h^{K,g}(\mathbf{i}) = f\mathbf{i}\}\right| \\
= \quad & \\
& \frac{1}{2^m} \left|\{\mathbf{i} \in 2^m \mid h^{K,g}(\mathbf{i}) \neq f\mathbf{i}\}\right| \\
= \quad & \{ \text{ definition of average distance, eq. (21) } \} \\
& \tilde{d}(h^{K,g}, f) \\
\geq \quad & \{ \text{ definition of } \tilde{\nu}, \text{ eq. (22), since } h^{K,g} \text{ is linear } \}
\end{aligned}$$

$$\tilde{\nu}(f) \ ,$$

which concludes the proof. □

E. Contextual fraction and games

We expand here on the last paragraph of the main text. We discuss an interpretation of the result of Theorem 1 in light of the logical description of Bell inequalities from [17]. This yields a result with a similar flavour to that of Theorem 3 relating the hardness of a task, its probability of success, and the contextual fraction of the resource used.

A constraint system is a specified by a tuple $\langle V, D, \Gamma\rangle$, where V is a finite set of variables, D a finite set called the domain, and Γ a finite set of propositional formulae with atoms of the form $(v = d)$ where v is a variable in V and d a value in D. We write $V(\phi)$ for the variables that appear in the formula $\phi \in \Gamma$.

We consider the following task. Given a formula $\phi \in \Gamma$ as input, one must reply with an assignment $s : V(\phi) \longrightarrow D$ of domain values to each variable that appears in the formula. Thus, a probabilistic strategy is a family $\{p_\phi\}_{\phi\in\Gamma}$ where p_ϕ is a probability distribution over $D^{V(\phi)}$. A strategy is considered valid if these probabilities satisfy the following compatibility condition:

$$\forall \phi_1, \phi_2 \in \Gamma,\ p_{\phi_1}|_{V(\phi_1)\cap V(\phi_2)} = p_{\phi_2}|_{V(\phi_1)\cap V(\phi_2)}$$

The idea is that, averaging over several runs, one should give consistent answers for the same variable apearing in different formulae. In specific cases where this game can be interpreted as a multi-player nonlocal game (see below), this condition corresponds to the imposed requirement of no communication between players.

The goal is to answer with an assignment $s : V(\phi) \longrightarrow D$ that satisfies the input formula ϕ (written $s \models \phi$, with the obvious meaning) as often as possible, while following a valid strategy. The **average probability of success** of a strategy $\{p_\phi\}_{\phi\in\Gamma}$ is given by

$$p_S := \frac{1}{|\Gamma|} \sum_{\phi\in\Gamma} \sum_{s\in D^{V(\phi)}, s\models\phi} p_\phi(s)$$

where we are assuming that all input formulae are equally probable.

Since the aim is to jointly satisfy a set of formulae, the notion of k-consistency is a measure of the hardness of the task. A set of formulae is said to be k-consistent if at most k can be jointly satisfied by an assignment of domain values to all the variables appearing in the formulae. If our set Γ of n formulae is k-consistent, the fraction $\frac{(n-k)}{n}$ is a normalised measure of the hardness of the task.

Note that all this is simply an alternative way of formulating measurement scenarios and empirical models. We can consider a scenario $\langle X, \mathcal{M}, O\rangle$ given by $X = V$, $O = D$, and with the contexts in $\mathcal{M}$ being the maximal sets of the form $V(\phi)$ with $\phi \in \Gamma$. Then, a valid strategy is simply a (no-signalling) empirical model for this scenario. As such, we can speak of its contextual fraction.

The particular case of Bell scenarios corresponds to a multi-player game where each player is responsible to answer for certain variables and where at most one variable from each player may appear in each formula $\phi \in \Gamma$. This is a nonlocal game as considered e.g. in [16, 32, 33], and there the validity condition on the strategy can be motivated as imposing no communication between players. In fact, there is a way of converting a general constraint satisfaction task into a two-player nonlocal game in such a way that the imposition of no-communication captures the strategy validity requirement. This transformation yields an interesting correspondence even at the level of quantum realisability [32–35], where it can be seen as establishing a tight connection between strong nonlocality and state-independent strong contextuality [35]. However, we shall not expand on this point here.

Combining our Theorem 1 with (a mild generalisation to arbitrary outcome sets of) the logical derivation of Bell inequalities from [17], we obtain the following result, whose form is analogous to that of Theorem 3.

Theorem 4. *Let $\langle V, D, \Gamma\rangle$ be a constraint system where Γ is k-consistent, and consider a valid strategy $\{p_\phi\}_{\phi\in\Gamma}$ with average probability of success p_S, and corresponding average probability of failure $p_F := 1 - p_S$. Then,*

$$p_F \geq \mathsf{NCF}(p)\frac{n-k}{k} \ .$$

Proof. Consider the inequality

$$\sum_{\phi\in\Gamma} \sum_{\substack{s\in D^{V(\phi)}\\ s\models\phi}} p_\phi(s) \leq k \ . \tag{25}$$

We first show that this is a Bell inequality, i.e. that it is satisfied by every non-contextual model. This amounts to a generalisation of [17, Proposition 4] to arbitrary outcome sets. It suffices to show that the inequality is satisfied by every non-contextual deterministic model (or strategy) – i.e. a model determined by some global assignment $t : V \longrightarrow D$ – since any non-contextual model is a convex combination of such models. Consider the (deterministic non-contextual) strategy p^t given by $p^t_\phi(s) = \delta_{s,t|_{V(\phi)}}$. Then

$$\sum_{\phi\in\Gamma} \sum_{\substack{s\in D^{V(\phi)}\\ s\models\phi}} p^t_\phi(s) = \sum_{\phi\in\Gamma} \sum_{\substack{s\in D^{V(\phi)}\\ s\models\phi}} \delta_{s,t|_{V(\phi)}} = \sum_{\substack{\phi\in\Gamma\\ t|_{V(\phi)}\models\phi}} 1 \leq k \ .$$

where the last step follows from the assumption that Γ is a k-consistent set of formulae.

We thus have a Bell inequality (25), whose algebraic bound is equal to $n = |\Gamma|$. Note that the left-hand side is equal to $n\,p_S$.

By Theorem 1, we must have that the normalised violation of this Bell inequality by (the empirical model corresponding to) the strategy $\{p_\phi\}_{\phi\in\Gamma}$ is at most $\mathsf{CF}(p)$:

$$\frac{\max\{0, n\,p_S - k\}}{n-k} \leq \mathsf{CF}(p) \ .$$

By substituting $(1-p_F)$ for p_S and $(1-\mathsf{NCF}(p))$ for $\mathsf{CF}(p)$ and simplifying, this easily seen to be equivalent to the desired relation. □