



**HAL**  
open science

## Corrected Bell and Noncontextuality Inequalities for Realistic Experiments

Kim Vallée, Pierre-Emmanuel Emeriau, Boris Bourdoncle, Adel Sohbi, Shane Mansfield, Damian Markham

### ► To cite this version:

Kim Vallée, Pierre-Emmanuel Emeriau, Boris Bourdoncle, Adel Sohbi, Shane Mansfield, et al.. Corrected Bell and Noncontextuality Inequalities for Realistic Experiments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 2023, 382 (2268), <10.1098/rsta.2023.0011>. <hal-04271961>

**HAL Id: hal-04271961**

**<https://hal.science/hal-04271961v1>**

Submitted on 6 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# Corrected Bell and Noncontextuality Inequalities for Realistic Experiments

Kim Vallée<sup>\*1</sup>, Pierre-Emmanuel Emeriau<sup>\*2</sup>, Boris Bourdoncle<sup>2</sup>, Adel Sohbi<sup>3</sup>,  
Shane Mansfield<sup>2</sup>, and Damian Markham<sup>1</sup>

<sup>1</sup>Sorbonne Université, CNRS, LIP6, F-75005 Paris, France

<sup>2</sup>Quandela, 7 Rue Léonard de Vinci, 91300 Massy, France

<sup>3</sup>ORCA Computing, Toronto, M6P 3T1, Canada

Contextuality is a feature of quantum correlations. It is crucial from a foundational perspective as a nonclassical phenomenon, and from an applied perspective as a resource for quantum advantage. It is commonly defined in terms of hidden variables, for which it forces a contradiction with the assumptions of parameter-independence and determinism. The former can be justified by the empirical property of non-signalling or non-disturbance, and the latter by the empirical property of measurement sharpness. However, in realistic experiments neither empirical property holds exactly, which leads to possible objections to contextuality as a form of nonclassicality, and potential vulnerabilities for supposed quantum advantages. We introduce measures to quantify both properties, and introduce quantified relaxations of the corresponding assumptions. We prove the continuity of a known measure of contextuality, the contextual fraction, which ensures its robustness to noise. We then bound the extent to which these relaxations can account for contextuality, via corrections terms to the contextual fraction (or to any noncontextuality inequality), culminating in a notion of genuine contextuality, which is robust to experimental imperfections. We then show that our result is general enough to apply or relate to a variety of established results and experimental setups.

## 1 Introduction

Contextuality is well-studied as a nonclassical feature of the empirical predictions of quantum mechanics, which may be considered as a generalisation of Bell nonlocality [1]. First studied by Bell [2] and Kochen and Specker [3, 4], there now exist a number of general, structural frameworks for treating contextuality, including the sheaf [5], graph [6], hypergraph [7], and contextuality-by-default [8] frameworks. The study of contextuality has a strong foundational interest, but also led to a range of recent results that establish links between contextuality and quantum advantage in the aforementioned frameworks [9, 10, 11, 12, 13, 14, 15, 16, 17]<sup>1</sup>, which makes it an important phenomenon for applications in quantum information and computation.

---

Kim Vallée\*: [kim.vallee@lip6.fr](mailto:kim.vallee@lip6.fr)

Pierre-Emmanuel Emeriau\*: [pe.emeriau@quandela.com](mailto:pe.emeriau@quandela.com)

<sup>1</sup>In particular, the randomness certification protocol demonstrated in [17], which is robust to cross-talk effects inherent to the experiment, makes use of techniques introduced in the present work.

As it is usually presented, the physical consequences of contextuality are such that for any hidden variable theory reproducing the observed behaviour it forces one to give up on either *determinism* or *parameter independence*. In this way, very roughly speaking, we could say that contextuality prevents the interpretation of measurements as revealing pre-determined values (determinism)—at least in a way that does not disturb the supposed underlying state of the system (parameter independence). The structural formalisms mentioned above admit equivalent definitions, but which shift the focus away from hidden variables, and bring into play new techniques and methodologies for analysing contextuality. However, in attempting to grasp the physical meaning of contextuality the hidden variable description is nevertheless often the most insightful. Through the article we will use a quantitative measure of contextuality, known as the Contextual Fraction (CF) [5], which has been shown to generalise Bell and noncontextual inequalities [11].<sup>2</sup>

However, in realistic, noisy experiments, the validity of the assumptions of *determinism* and *parameter independence* can be called into question. As we will discuss, and as has been pointed out elsewhere (e.g. [20]), beyond certain noise thresholds a contextual experiment can admit explanations in line with classical intuitions. While noise does not directly feature in hidden variable models, a standard justification for hidden variable assumptions is that they reflect some empirical property [21]. So if noise affects those properties it will also weaken the validity of the assumptions.

In the case of *determinism*, the analogous empirical property is *sharpness* (usually of measurements). In theory at least, classical and quantum physics allow for measurements that are perfectly sharp: e.g. if the measurement is immediately repeated it returns the same outcome [22]. An extreme example of a dichotomic unsharp measurement would be one which independent of the state returns outcomes 0 or 1 with equal probability.

While this is perfectly reasonable as a measurement, it would be unreasonable to assume that such a measurement is capable of revealing pre-determined values of some properties of the state it is applied to. In practice, realistic measurements may be sharp to a good approximation but will never be perfectly sharp.

In the case of *parameter independence* the analogous empirical property is *no-signalling* or *no-disturbance*. In Bell-type scenarios, no-signalling is the property that any party's measurement choice does not influence other parties' measurement outcomes. In contextuality scenarios like those of [5, 6, 7, 8], we do not need to assume different spacelike separated parties but may still speak of a generalised no-signalling property sometimes referred to as *no-disturbance*. In theory at least, the empirical behaviours of systems that obey classical or quantum physics should be perfectly non-signalling in this generalised sense. In practice, however, no realistic empirical model can ever be *perfectly* non-signalling. This could simply be due to finite statistics, or have to do with systematic errors such as imperfect shielding or experimental crosstalk. Our work is concerned with systematic errors and the properties of the behaviour underlying given empirical data. For methods to deal with the problem of finite statistics in experimental trials, we refer the reader to [23, 24] when testing local realism, and to [25, 26] for device-independent information processing.

So if we never observe perfectly sharp nor non-signalling empirical data, why should we expect an underlying hidden variable or physical description of the system be perfectly deterministic and parameter independent? We propose to quantify unsharpness and signalling, and to admit hidden variable models that are nondeterministic and parameter dependent to a corresponding extent. Contextuality should only be considered robust to

---

<sup>2</sup>The CF can be viewed as a generalisation of the non-local content defined in [18]. A related notion of noncontextual content [19], which appeared later, corresponds to  $1 - \text{CF}$ ; in other words it is the noncontextual fraction.

Empirical property	Hidden-variable property
No-Signalling (NS)	Parameter Independence (PI)
Sharpness	Outcome Determinism (OD)

Table 1: Comparison between characteristics of the empirical data and its counterpart at the hidden-variable level.

the above objections if it cannot be accounted for by this additional flexibility afforded to hidden variable explanations. In other words, any attempt to explain robust contextuality would necessitate *more* non-determinism or *more* disturbance than is empirically justified.

**Relationship to Contextuality-by-Default (CbD)** The CbD approach to contextuality [8] is a general (hidden-variable) model-independent treatment of contextuality in the presence of signalling, in part motivated by applications to empirical data arising both in physical [27] and non-physical settings, e.g. in linguistics [28, 29] or psychology [30, 31]. A key component of CbD analysis is the notion of a maximal coupling of empirical data. Although it is not the usual perspective of CbD, maximal couplings can be viewed as a kind of hidden variable model in which the amount of parameter dependence is constrained to match the amount of signalling present in the empirical data, wherever measurements appear in multiple contexts. At a conceptual level there is thus a clear similarity between CbD and the present approach. There are also some clear distinctions. We wish to reason about which physical mechanisms could underlie a given empirical behaviour. For that reason, from the outset, we explicitly anchor the foundations of our approach in terms of hidden variable models, as they provide a broad framework for hypothesising and reasoning about such mechanisms. By design, it is therefore clearer how to extract a physical meaning from an observation of genuine contextuality with our approach.<sup>3</sup>In contrast to CbD, the present approach admits the degree of parameter dependence to be bounded by empirical signalling in a flexible manner – in particular without requiring the bounds to be saturated, allowing for different kinds of hidden variable explanation. In any case the quantification of signalling and parameter dependence in both approaches are different. Our quantification methods are motivated by natural connection or analogy to the contextual fraction [11]. This is known to coincide [32, 33] in some specific scenarios with measures used within the CbD approach [34], which however are only defined for dichotomic measurements. Finally, while the CbD approach applies to signalling, we will also bring into play considerations of sharpness.

**Relationship to other previous works** Previous analyses have focused specifically on robustness to unsharpness [35, 36, 37], or to signalling [38, 8]. Others have considered generalised notions of hidden variable models that allow extra expressivity through state updating [39, 40]. The work of [41] takes into account both signalling and unsharpness, although the analysis is limited to the Bell-CHSH scenario; in particular our approach retrieves those results and extends them to any scenario.

Our aim is to provide a simplified, holistic approach that is robust to both objections. We show that experiment-friendly [17] inequalities that straightforwardly extend the contextual fraction measure [11] and test for witnessing *genuine* contextuality can be obtained

---

<sup>3</sup>The approach may have significance for reasoning about mechanisms underlying empirical data in other settings too, but here our primary focus is on their well-established application to physical phenomena.

by setting lower bounds on the admissible fraction of nondeterminism  $\eta$  and parameter-dependence  $\sigma$ .

## 2 General framework for contextuality

Here we briefly summarise some of the main ideas from [5, 11]. An extended introduction can be found in the Supplemental Material.

### 2.1 Ingredients à la Abramsky-Brandenburger

*Measurement scenario.*— An abstract description of an experimental setup is described by a measurement scenario  $\langle X, \mathcal{M}, O \rangle$  where (i)  $X$  is a finite set of measurement labels; (ii)  $\mathcal{M} \subseteq \mathcal{P}(X)$  is the set of all maximal contexts, i.e. sets of compatible measurements, and forms a cover of  $X$ ; (iii)  $O = (O_x)_{x \in X}$  where  $O_x$  is the outcome set corresponding to measurement  $x \in X$ .

*Empirical behaviours.*— Given the description of the experimental setup, either calculating theoretical predictions for admissible joint outcomes or performing repeated runs of the experiment with varying choices of measurement context and recording the frequencies of the corresponding joint events results in a probability table which is formalised as an empirical model. Formally, an empirical model, or behaviour,  $e$  on  $\langle X, \mathcal{M}, O \rangle$  is a family  $e = (e_C)_{C \in \mathcal{M}}$  where  $e_C$  is a probability distribution on the corresponding joint outcome space.

*Non-signalling.*— A behaviour  $e$  is non-signalling when for any two contexts  $C_1$  and  $C_2$ ,  $e_{C_1}|_{C_1 \cap C_2} = e_{C_2}|_{C_1 \cap C_2}$ . The notation  $e_C|_U$  with stands for the marginalisation of the probability distribution  $e_C$  to  $U \subseteq C$ : for  $t \in O_U$ ,  $e_C|_U(t) := \sum_{s \in O_C, s|_U=t} e_C(s)$ , where  $s|_U$  is simply the function restriction of  $s$  to the domain  $U$  and  $O_U := \prod_{x \in U} O_x$ .

*(Non)contextuality.*— Informally, an empirical behaviour  $e$  is said to be noncontextual whenever context-wise predictions can all be obtained as the marginals of one probability distribution over global value assignments.

Note at this point that the statement concerns only the empirical model, and we have not referred to any underlying hidden variable model. The relationship with hidden variable models was made explicit in [5] for the case of perfectly parameter-independent hidden variable models (see Proposition 1). Thus the Abramsky–Brandenburger approach is equivalent to, yet neatly abstracts away from, hidden-variable models. It also leads to convenient tools like the contextual fraction to quantify contextuality [11], which we recap next. Establishing an analogous relationship for noisy data, and a convenient method for quantifying contextuality in a noise-robust manner contextual is the main contribution of this work.

### 2.2 Quantifying contextuality and signalling

By definition  $O_X$  is the set of global assignments of an outcome to each measurement. Let  $n := |O_X|$  be the number of global assignments and  $m := \sum_{C \in \mathcal{M}} |O_C|$  the number of total ‘local’ (or context-wise) assignments ranging over all contexts. We use bold notation for vectors. Local assignments can be listed as:  $\{\langle C, \mathbf{s} \rangle \text{ s.t. } C \in \mathcal{M} \text{ and } \mathbf{s} \in O_C\}$ . Then the incidence matrix  $M$  which records the possible restrictions from global assignments  $g$  to local assignments  $\langle C, \mathbf{s} \rangle$  is a  $m \times n$  (0,1)-matrix defined as:

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

To understand its action, consider the columns of  $M$  as listing the global assignments  $\mathbf{g} \in O_X$ . In particular, notice that rows are labelled by local assignments  $\mathbf{s}$ , and that for any given row and column,  $M$  assigns the value 1 whenever the corresponding local and global assignments agree,  $\mathbf{g}|_C = \mathbf{s}$  (and 0 otherwise).

The empirical model  $e$  can be represented as a vector  $\mathbf{v}^e \in [0, 1]^m$  where for a given context  $C \in \mathcal{M}$  and a local assignment  $\mathbf{s} \in \mathcal{O}_C$ ,  $\mathbf{v}^e[\langle C, \mathbf{s} \rangle] = e_C(\mathbf{s})$ . This is the flattened version of the tables that are usually used to represent empirical models.

*Contextual fraction.*— Given empirical behaviours  $e_1$  and  $e_2$  on  $\langle X, \mathcal{M}, O \rangle$ ,  $e = \lambda e_1 + (1 - \lambda)e_2$  for  $\lambda \in [0, 1]$  is another valid empirical model. Instead of asking whether  $e$  is contextual, we can wonder what fraction of  $e$  can be explained noncontextually. This amounts to looking for a convex decomposition of the form  $e = \lambda e^{\text{NC}} + (1 - \lambda)e'$  with weight  $\lambda$  on the noncontextual part. Maximising  $\lambda$  yields the irreducible part of contextuality called the contextual fraction (see Figure 1):

$$\begin{aligned} \text{CF}(e) &:= 1 - \text{NCF}(e) \\ &:= 1 - \max\{\lambda \mid \exists e^{\text{NC}} \text{ non-contextual for which } e' = \lambda e^{\text{NC}} + (1 - \lambda)e'\}. \end{aligned} \quad (2)$$

It has a number of desirable properties [11] including that it can be computed as a linear program:

$$\text{(P-NCF)} \quad \left\{ \begin{array}{l} \text{Find } \mathbf{b} \in \mathbb{R}^n \\ \text{maximising } \mathbf{1} \cdot \mathbf{b} \\ \text{subject to:} \\ \quad \mathbf{M}\mathbf{b} \leq \mathbf{v}^e \\ \quad \mathbf{b} \geq \mathbf{0} . \end{array} \right.$$

This program computes the noncontextual fraction  $\text{NCF}(e)$  of  $e$ . We refer the interested readers to the Appendices for more details.

*Signalling fraction.*— In the same spirit, we introduce the signalling fraction  $\text{SF}(e)$  of a signalling behaviour  $e$  by looking for a convex decomposition within the signalling polytope (see Figure 1) that maximises the weight of the non-signalling part:

$$\begin{aligned} \text{SF}(e) &:= 1 - \text{NSF}(e) \\ &:= 1 - \max\{\lambda \mid \exists e^{\text{NS}} \text{ non-signalling for which } e' = \lambda e^{\text{NS}} + (1 - \lambda)e'\}. \end{aligned}$$

This maximised weight can also take values in the interval  $[0, 1]$  and is called the signalling fraction<sup>4</sup>. Just like the contextual fraction, the irreducible weight on the signalling part is the signalling fraction and can be computed with a linear program. Idealised empirical behaviours predicted by quantum mechanics, in which contexts consist of perfectly commuting observables, will not be signalling. However, in realistic settings noise can manifest itself to introduce some amount of signalling and the signalling fraction defined above will quantify this as a property of the empirical data.

Just like the contextual fraction, the signalling fraction of any empirical behaviour can

---

<sup>4</sup>The signalling fraction originates from unpublished work of Samson Abramsky, Rui Soares Barbosa and SM.

be efficiently computed through a linear program:

$$(P\text{-NSF}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{b} \in \mathbb{R}^n \\ \text{maximising } \mathbf{1} \cdot \mathbf{b} \\ \text{subject to:} \\ \quad \mathbf{M}\mathbf{b} \leq \mathbf{v}^e \\ \quad \mathbf{M}\mathbf{b} \geq 0 . \end{array} \right.$$

Again this program is computing the no-signalling fraction  $\text{NSF}(e) := 1 - \text{SF}(e)$  of  $e$ . Further details can be found in the Appendices.

### 2.3 Relaxing parameter independence and outcome determinism

*Hidden variable models.*— Hidden variable models (HVMs) provide a broad, general approach to considering possible physical explanations underlying observed data<sup>5</sup>. We usually place assumptions on HVMs to reflect notions of classicality—for instance outcome determinism and parameter independence [5].

Formally a HVM on  $\langle X, \mathcal{M}, O \rangle$  is a triple  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  where: (i)  $\Lambda$  is a finite space of hidden variables; (ii)  $p$  is a probability distribution on  $\Lambda$ ; (iii) for each  $\lambda \in \Lambda$ ,  $h^\lambda$  is a behaviour on  $\langle X, \mathcal{M}, O \rangle$ .

Essentially each hidden variable has a behaviour associated with it. We may only have probabilistic information about which hidden variable pertains, reflected by the probability distribution  $p$ . This gives rise to an effective overall empirical behaviour  $h := \sum_{\lambda \in \Lambda} p(\lambda)h^\lambda$ . A sufficient condition for  $h$  to be *non-signalling* is *parameter independence*, which requires that for all  $\lambda \in \Lambda$  and contexts  $C_1, C_2 \in \mathcal{M}$ ,  $h_{C_1 \cap C_2}^\lambda = h_{C_2}^\lambda|_{C_1 \cap C_2}$ .

To ensure *sharpness* we can impose *outcome determinism* i.e. that  $\forall \lambda, \forall C \in \mathcal{M}$ ,  $h_C^\lambda$  is a Dirac distribution on a given joint outcome in  $O_C$ . Note that, in contrast to parameter independence, it is always possible to find a decomposition of an empirical behaviour into deterministic hidden variables (which may be signalling). The following proposition is a corollary of [5, Prop. 3.1 and Th. 8.1].

**Proposition 1** (from [5]). *An empirical behaviour  $e$  is noncontextual ( $\text{CF}(e) = 0$ ) if and only if it is realisable by a parameter-independent, deterministic HVM.*

We relax the assumptions of OD and PI below.

**Definition 1** ( $(1 - \sigma)$  PI HVM). *A hidden-variable model is said to be  $(1 - \sigma)$  parameter-independent if for all hidden variables  $\lambda$ , there exists a decomposition of  $h^\lambda$  of the form  $h^\lambda = (1 - \sigma_\lambda)h_{NS}^\lambda + \sigma_\lambda h'^\lambda$  with  $h_{NS}^\lambda$  a parameter-independent model and with necessarily  $\sigma_\lambda \leq \sigma$ .*

Note that  $\sigma$  is not necessarily the signalling fraction of the overall empirical behaviour realised by the HVM, but rather the maximum parameter dependence allowed for each hidden variable behaviour.

**Definition 2** ( $(1 - \eta)$  OD HVM). *A hidden-variable model is said to be  $(1 - \eta)$  outcome deterministic if for all hidden variables  $\lambda$ , there exists a decomposition of  $h^\lambda$  of the form  $h^\lambda = (1 - \eta_\lambda)h_{OD}^\lambda + \eta_\lambda h''^\lambda$  with  $h_{OD}^\lambda$  an outcome deterministic behaviour (which may be parameter-dependent) and  $\eta_\lambda \leq \eta$ .*

---

<sup>5</sup>Sometimes HVMs may also be referred to as ontological models – attempts to capture an underlying ontology or physical reality. Note however that in some circumstances this terminology may implicitly assume additional structure to what define for HVMs here [42].

An interesting and well-studied behaviour is the PR-box [43]. It is a non-signalling, maximally contextual ( $\text{CF} = 1$ ) empirical model. In other words it does not admit a HVM realisation that satisfies the strict definitions of parameter independence and outcome determinism. However, under our relaxed definitions, we could in fact explain PR-box correlations by HVMs by allowing either (i) a degree of outcome non-determinism or (ii) parameter dependence (or a combination of both).

To see how PR-box behaviour can be explained by outcome non-deterministic HVMs, we assume parameter independence ( $\sigma = 0$ ) and note that it only admits a  $(1 - \eta)$  outcome-deterministic HVM with  $\eta = 0.5$ . To see this, the model is composed of a single hidden variable behaviour, the PR-box itself, which is parameter-independent:

$$h_{\text{PR}} = h_{\text{PR}}^\lambda. \quad (3)$$

Since it's possible to express the hidden variable behaviour of the PR-box as a mixture of two deterministic (but parameter-dependent) behaviours:

$$h_{\text{PR}}^\lambda = \frac{1}{2}h_{M_1}^\lambda + \frac{1}{2}h_{M_2}^\lambda, \quad (4)$$

then, by Definition 2,  $\eta = 1/2$ . On the other hand the PR-box could also arise from a purely outcome deterministic ( $\eta = 0$ ) HVM. Revisiting the decomposition from Eq. (4), this can be seen as describing the PR-box by an outcome deterministic hidden variable model containing two maximally signalling hidden variable behaviours ( $\sigma = 1$ ) which each pertains with equal probability:  $e^{\text{PR}} = \frac{1}{2}h_{M_1}^{\lambda_1} + \frac{1}{2}h_{M_2}^{\lambda_2}$ . For a longer discussion on the PR-box we refer the reader to appendix (A.d).

Note that the assumptions of determinism and parameter independence are not quite on an equal footing. If parameter independence is relaxed completely (that is, allowing  $\sigma = 1$ ) then determinism becomes trivial in the sense that any behaviour will admit deterministic HVMs. On the other hand, if determinism is relaxed, but parameter independence is respected, then only a non-signalling behaviour may be constructed. A related discussion can be found in [44].

### 3 Continuity of the contextual fraction

Since realistic experiments are inherently noisy, it is important that any measure of contextuality is robust to noise. At the very least it should satisfy some notion of continuity with respect to some total variation distance  $V$  generalised to empirical models (see Supplemental Material). The main theorem in this section is thus significant independently of the rest of the present work, and adds to the list of desirable properties of the contextual fraction measure identified in [11].

**Theorem 1.** *Let  $e$  and  $e'$  be empirical models on the measurement scenario  $\langle X, \mathcal{M}, O \rangle$ . If  $V(e, e') \leq \varepsilon$  for  $\varepsilon > 0$  then  $|\text{CF}(e) - \text{CF}(e')| \leq |\mathcal{M}|\varepsilon$ .*

A detailed proof is given in Supplemental Material.

*Sketch of proof.* We use the theory of finite dimensional linear programming, and more precisely perturbation theory [45, Section 5.6]. The idea is to control perturbation of the optimal value of the primal program by exhibiting optimal solutions of the dual program and a perturbed dual program.  $\square$

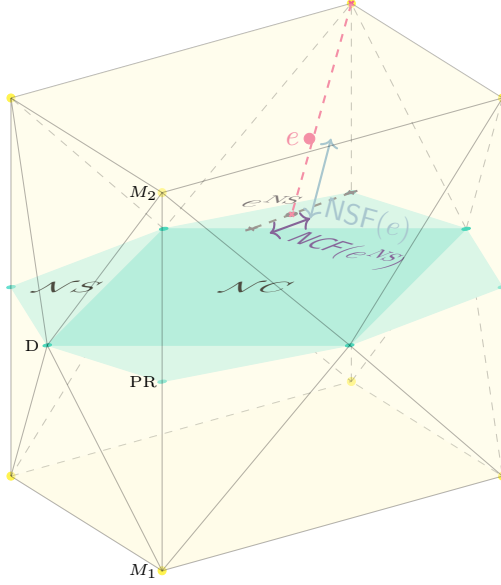


Figure 1: A schematic representation of a cut of the signalling correlation polytope for the (2,2,2) Bell scenario with the no-signalling (NS) polytope and the noncontextual (NC) polytope inscribed in it. Vertices of the polytopes are highlighted: dark blue vertices correspond to NC sharp behaviours; light blue vertices to maximally contextual yet no-signalling behaviours (forming the NS polytope); and yellow vertices correspond to the signalling sharp behaviours. Note that contextual vertices of the non-signalling polytope, unlike noncontextual ones, are not vertices of the signalling polytope. The quantum set (not represented) in general corresponds to a convex set between the NS and NC polytopes. For a signalling behaviour  $e$ , its non-signalling fraction (maximal weight on the non-signalling part  $e^{\text{NS}}$  in a convex decomposition) is represented. Likewise, the noncontextual fraction of  $e^{\text{NS}}$  is represented. Labels displayed for some vertices are to be referred to in Figure 2.

## 4 Corrected Bell and Noncontextuality Inequalities

The nonclassicality criterion  $\text{CF} > 0$  applies for perfectly sharp and no-signalling behaviours. Below we want to relax these assumptions and still derive an inequality that captures genuine nonclassicality.

### 4.1 Relaxing parameter independence and sharpness

**Lemma 1.** *Let  $\langle \{\lambda\}, \delta_\lambda, (h^\lambda) \rangle$  be a  $(1-\eta_\lambda)$  OD and  $(1-\sigma_\lambda)$  PI HVM. Then, if  $2\eta_\lambda + \sigma_\lambda < 1$ , the contextual fraction of the hidden variable model satisfies  $\text{CF}(h^\lambda) \leq \eta_\lambda$ .*

A detailed proof is given in Supplemental Material.

*Sketch of proof.* Using Definition 2, we can distinguish two cases for an optimal outcome deterministic decomposition: either  $h_{\text{OD}}^\lambda$  is parameter-independent, in which case we compare with (2) to see that the bound is respected, or  $h_{\text{OD}}^\lambda$  is parameter-dependent, in which case we show that  $\sigma_\lambda + 2\eta_\lambda \geq 1$  contradicting the initial assumptions.  $\square$

**Theorem 2.** *Let  $e = \sum_\lambda p(\lambda)h^\lambda$  be a behaviour realisable by a  $(1-\sigma)$  PI and  $(1-\eta)$  OD HVM  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  such that  $\sigma + 2\eta < 1$ . Then its contextual fraction is bounded above by:  $\text{CF}(e) \leq \eta$ .*

A detailed proof is given in Supplemental Material.

*Sketch of proof.* By applying Lemma 1 to all hidden variable behaviours and using the convexity of the contextual fraction, we obtain Theorem 2.  $\square$

A similar result was obtained previously in [41] specifically for the Bell-CHSH scenario, with similar definitions of parameter dependence and determinism. In comparing the works, one should note that here we assume measurement independence (also related to experimenter free-will), that is  $p(\lambda|x, y) = p(\lambda|x', y') = p(\lambda)$  and thus don't include correction terms related to measurement dependence in our inequality.

The condition  $2\eta_\lambda + \sigma_\lambda < 1$  from Lemma 1 can be motivated either from experimental considerations or by taking a more theoretical approach. First, in real experimental setups, we expect a low-noise regime so that we remain close to the quantum (or classical) set. In this case the ontological conditions  $2\eta_\lambda + \sigma_\lambda < 1$  should be respected. Secondly, for  $n$ -cycle scenarios [46], one can find a hidden variable behaviours which respect  $2\eta_\lambda + \sigma_\lambda = 1$  such that  $\text{CF}(h^\lambda) = 1$ . The proof is given in the Supplemental Material and Figure 2 gives a geometric intuition for why this holds. In this case, the inequality for genuine nonclassicality ( $\text{CF} > 1$ ) would become trivially unsatisfiable. For more general scenarios it is an open problem to know whether there always exists at least one hidden variable behaviour such that  $\text{CF} = 1$  when  $2\eta_\lambda + \sigma_\lambda \geq 1$ .

We note that in the case of determinism ( $\eta = 0$ ), due to Definitions 1 and 2, we have either  $\sigma = 0$  or  $\sigma = 1$ . Due to the condition  $\sigma + 2\eta < 1$ , Theorem 2 can either be applied for a parameter-dependent HVM or it cannot be applied since  $\sigma = 1$ . Instead, this result is of most use when parameter dependence can be accounted for by a lack of determinism:  $\eta \geq \sigma$ .

## 4.2 Relaxing only parameter independence

As noted earlier, parameter dependence and determinism are not on an equal footing. For instance relaxing completely parameter independence trivialises outcome determinism. We note that our previous derivation in Section 4.1 is quite rigid for determinism. Following Definition 1, if we impose outcome determinism then the following holds:

$$\eta = 0 \implies \sigma \in \{0, 1\}. \quad (5)$$

This implies that if one has a perfectly sharp experiment, then parameter dependence is either forbidden  $\sigma = 0$  or always maximally relaxed  $\sigma = 1$ .

Let us review another way of looking at only relaxing parameter-independence while preserving outcome determinism. Another form of relaxation has been used in various references e.g. [47, 38]. These relaxations have the assumption of outcome determinism ( $\eta = 0$ ), defined in the same manner as Definition 2 but they relate parameter dependence to another quantity that we will call  $\sigma'$  here. Generally speaking, in the case of determinism, each hidden variable behaviour  $h^\lambda$  is either parameter-independent or maximally parameter-dependent. Thus we can split the sum of all hidden variables depending on their parameter-dependence:

$$e = \sum_{h^\lambda \in \text{PI}} p(\lambda)h^\lambda + \sum_{h^\lambda \notin \text{PI}} p(\lambda)h^\lambda, \quad (6)$$

where PI here is used as a shorthand for the set of hidden variable behaviours that are parameter-independent.

The new definition of parameter independence is now a bound on the probability to have parameter-dependent hidden variables by  $\sigma'$ :

$$\sum_{h^\lambda \notin \text{PI}} p(\lambda) \leq \sigma'. \quad (7)$$

This in turn leads easily to a bound on the contextual fraction (due to its convexity):

$$\begin{aligned}
\text{CF}(e) &\leq \sum_{h^\lambda \in \text{PI}} p(\lambda) \text{CF}(h^\lambda) + \sum_{h^\lambda \notin \text{PI}} p(\lambda) \text{CF}(h^\lambda) \\
\text{CF}(e) &\leq \sum_{h^\lambda \notin \text{PI}} p(\lambda) \text{CF}(h^\lambda) \\
\text{CF}(e) &\leq \sum_{h^\lambda \notin \text{PI}} p(\lambda) \\
\text{CF}(e) &\leq \sigma'.
\end{aligned} \tag{8}$$

Note again that this derivation only holds for deterministic hidden variables.

## 5 Approaches to experimental bounds

To relate Lemma 1 to statistics gathered during experiments, it is necessary to either physically impose restrictions on the values of  $\sigma$  or  $\eta$  or to make assumptions on their bounds based on observations from experiments. In Bell scenarios, for example, separation of parties allows us to make the assumption that the HVM should be parameter-independent i.e.  $\sigma = 0$ . In this case, Fine's theorem [48] (more precisely its generalisation in [5]) states that outcome determinism ( $\eta = 0$ ) is equivalent to non-contextuality ( $\text{CF} = 0$ ). However in the absence of spacelike separation ( $\sigma > 0$ ), we need to make assumptions on how experimental data is related to possible HVMs.

**Parameter-independence.** By definition,  $\sigma \geq \text{SF}(e)$  for any HVM realisation of a behaviour  $e$ . It means that the allowed parameter-dependent fraction  $\sigma$  cannot be lower than the observed signalling fraction of the behaviour  $e$  since  $\text{SF}$  is the optimal solution of a minimisation program. In a general manner, one wishes however to be able to upper bound the parameter dependence of the hidden variables according to the observed signalling:

$$\sigma \leq f(\text{SF}(e)) ,$$

where the function  $f$  is to be chosen, perhaps motivated by experimental considerations, through partial device characterisation, space-like shielding, or any other appropriate method. The simplest assumption would thus be  $\sigma = \text{SF}(e)$  but it is important to recall that whereas parameter independence of a HVM implies that the empirical behaviour it realises is non-signalling, the converse does not hold in general, so this is a strong assumption to impose. Note, however, that the analogous assumption is made in the contextuality-by-default approach to contextuality which also applies in the presence of signalling, and has been employed in experimental tests of contextuality analysed via contextuality-by-default [27] and the present approach [17].

**Outcome determinism** For determinism, an analogous approach would be to bound  $\eta$  by an appropriate quantification of the unsharpness of the empirical measurements. A convenient approach could be to follow the definition of sharpness found in [22]. Each context of measurements could be performed not once, but twice in quick succession, and the probabilities of seeing a change in outcomes between successive measurements would give an indication of the unsharpness. This does however entail an additional assumption of measurement *non-invasiveness* [49, 50], i.e. that measurement does not alter the underlying state of the system (neither the value of  $\lambda$  nor the response  $h^\lambda$ ). In this way it allows for effectively accessing properties of  $h^\lambda$ . This can be a sensible approach in experiments such

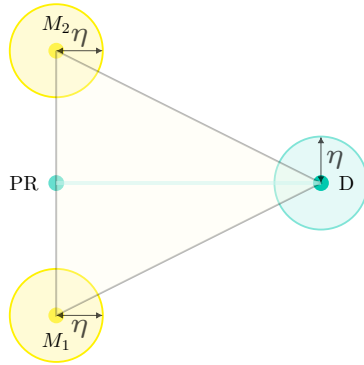


Figure 2: A schematic diagram of a vertical cut of Figure 1. Here, discs of radius  $\eta$  have been added around the deterministic points to represent the accessible hidden variable behaviours for a given  $\eta$ . Theorem 2 applied in this scenario means that whenever  $\sigma + 2\eta < 1$ , the only accessible behaviours are the ones within the light blue disc surrounding the deterministic classical blue point. On the other hand, when  $\sigma + 2\eta \geq 1$  the behaviours inside the intersections of the yellow discs (centred on maximally signalling vertices) and the signalling polytope become accessible, and thus  $\text{CF} = 1$  is reachable.

as that of Kirchmair *et al.* [39, 47] in which sequences of measurements were performed on trapped ions.

For cases where repeated measurements of a system are not practical, for example in optical setups where photons are absorbed during the measurement, we must follow a different approach. One possibility is to assume the existence of a particular set of preparations that allows us to assess if a measurement device is responding deterministically or not (similar to the assumption of a tomographically complete set of preparations used in [51, 20, 37, 36]). This is motivated by the fact that in the ideal quantum mechanical description for any sharp measurement there is always a set of states which return deterministic responses. If we assume that we have such special, preparable states, they can be used to quantify outcome determinism.

## 6 Relationship to previous results

We now discuss how our results relate to others that have previously appeared in the literature [47, 38, 52, 53, 54]. To this end we will consider how to fine-tune the parameters appearing in Theorem 2 or the derivation in Section 4.2 to adjust it to some specific experimental setups or data. In particular we will show how several previous results are contained and can be extended within our approach. In this section our intention is to be illustrative rather than exhaustive with respect to existing literature.

### 6.1 Relaxing determinism in known frameworks

As already discussed, in the case of determinism, we could apply directly the derivation from Section 4.2. In this case we would obtain directly the results from [47, 38]. However, one may also be interested in relaxing determinism and looking at how our result obtained in Theorem 2 could be apply in this setting. In the following, we thus aim to relate Theorem 2 to results imposing determinism.

**Ref. [47].** We will show that relaxing determinism is an alternative to relaxing parameter dependence for a high enough relaxation. Reference [47] is concerned with the order in which measurements are performed within a context but this is not something we have sought to capture in our approach. Our approach may still be applied, but will simply

treat such orderings as the same context. The correction term in [47] arises as a sum over probabilities that the outcome of an observable ‘flips’ in a given context. Roughly:

$$\varepsilon = \sum_{C \in \mathcal{M}} p^{\text{flip}}[C],$$

where  $p^{\text{flip}}[C]$  is the probability that at least one measurement outcome in context  $C$  is flipping. As an example, in the CHSH scenario where  $C = AB$  we have  $p^{\text{flip}}[AB] = p[(a = 1|A) \text{ and } (a = 0|AB)] + p[(a = 0|A) \text{ and } (a = 1|AB)]$  where  $a$  is the outcome of  $A$ . In the CHSH scenario, the bound becomes <sup>6</sup>:

$$\begin{aligned} \text{CF}(e) &\leq p^{\text{flip}}[AB] + p^{\text{flip}}[AB'] + p^{\text{flip}}[A'B] + p^{\text{flip}}[A'B'] \\ &\leq \varepsilon, \end{aligned} \tag{9}$$

We propose instead to consider the case where the flipping occurs only due to lack of determinism. We thus relax determinism by making the assumption that signalling observed at the empirical level is the result of non-determinism only:

$$\eta := \max_{C \in \mathcal{M}} (p^{\text{flip}}[C]). \tag{10}$$

In this case, we are in the regime where  $\eta \geq \sigma$ . If  $2\eta + \sigma < 1$ , then we can simply apply Theorem 2:

$$\text{CF}(e) \leq \max_{C \in \mathcal{M}} (p^{\text{flip}}[C]). \tag{11}$$

**Ref. [38].** This work also aims at correcting inequalities with an additional term depending on

$$\Pr(o_x^C \neq o_x^{C'}) \leq \varepsilon \tag{12}$$

where  $o_x^C$  is the outcome of some observable  $x$  in context  $C$ , and  $C \cap C' \neq \emptyset$ . The corrected bound proposed by [38] is the following:

$$\text{CF}(e) \leq \frac{\sum_i \omega_i (k_i - 1)}{\beta_{max} - \beta_{cl}} \varepsilon, \tag{13}$$

where  $\omega_i$  are the weights of the inequality being tested,  $\beta_{max}$  is its algebraic bound and  $\beta_{cl}$  its classical bound, and  $k_i$  is the degree of measurement  $i$  (the number of times it appears in different contexts, e.g.  $k_i = 2$  for all measurements in the CHSH scenario). In the same spirit as our approach to Ref. [47], we assume that the difference of the outcomes in different context is just due to lack of determinism:

$$\Pr(o_x^C \neq o_x^{C'}) \leq \eta. \tag{14}$$

Which, by application of Theorem 2, gives:

$$\text{CF}(e) \leq \eta. \tag{15}$$

We now compare the two bounds in different known scenarios.

Following [38], we are interested in ranges of  $\varepsilon$  for which this bound on the contextual fraction is less than or equal to  $\text{CF}_Q$ , the maximum attainable contextual fraction over

---

<sup>6</sup>Note that this is an upper bound, since it could be that the flipping of  $A$  and  $B$  always occur together, which means we would double the impact of the flipping on the violation.

quantum mechanical models. Otherwise noise would be too high to tell apart quantum behaviours from noisy classical ones. While there is no general relation between  $\sum_i \omega_i(k_i - 1)$  and  $\beta_{max} - \beta_{cl}$  in Equation (13), in practical examples note that the numerator is much larger than the denominator. We can take as an example the Peres-Mermin square [55, 39]. In our approach saturating the quantum bound, which for the Peres-Mermin square coincides with the algebraic bound, would require  $\eta = 1$ , whereas for [38] it is saturated once  $\varepsilon = \frac{6-5}{72} \approx 0.0138$ . Thus for the same ontological assumptions, we obtain a non-classicality witness which remains useful in a much higher noise regime.

## 6.2 Application to experimental results

This subsection considers experimental works that did not apply corrected inequalities, or to which we wish to directly compare the result of Theorem 2. Unlike in the previous section, we directly take the values from the experiments to compute the corrected bound. This subsection may serve as an example of various means to compute  $\eta$  and  $\sigma$  from experiments, by making assumptions on the hidden variables from the empirical behaviour.

**Ref. [54].** The authors reproduce a Hardy-like KCBS inequality, such that in theory  $p(1, 1|i, i + 1) = 0$  for  $i \in \{1, 2, 3, 4, 5\}$  for sharp measurements. Yet in practice they do obtain some events  $(1, 1|i, i + 1)$ , which means that part of the process is noisy (whether it is preparation, measurement or transformation). It makes sense to assume then that measurements are unsharp and the sharpness of measurement can be a good indicator of the hidden variable properties. If the detector fails  $\eta$  of the time then the HV behaviours would be  $1 - \eta$  of the time deterministic (same detectors click) and  $\eta$  of the time giving a random result. Thus we propose to chose

$$\eta = \max_i (p(1, 1|i, i + 1)) \quad (16)$$

such that the unsharpness reflects the value of  $\eta$ . With this choice we retrieve the same result as that of [54].

The assumption made in the aforementioned reference is that  $\varepsilon$  of the time the maximum violation is attained, leading to the following corrected inequality :

$$|\langle KCBS \rangle| \leq (1 - \varepsilon)\beta_{cl} + \varepsilon\beta_{max}, \quad (17)$$

which in terms of CF is equivalent to

$$CF \leq \varepsilon. \quad (18)$$

Our result—derived from ontological considerations—states that the maximum non-determinism is given by  $\eta$ , leading to the same mathematical conclusion:

$$CF \leq \eta. \quad (19)$$

Since we arrive at the same mathematical conclusion, we can tell that their violation is robust to the amount of observed unsharpness. They get  $\eta = 0.021$  while they obtain a violation  $CF \approx 0.16$  which is largely above the expected bound given by Theorem 2.

**Ref. [52].** Our work may be applied to the experimental procedure described in [52] to test the KCBS inequality [56]. The experimental setup is designed to ensure that measurements cannot have context dependence, up to a set of assumptions, except for one measurement  $A_1$  appearing in contexts  $A_1A_2$  and  $A_5A_1$ . They thus label the measurement appearing in these contexts as  $A_1$  and  $A'_1$  respectively, so that the modified KCBS inequality looks like:

$$\langle A_1A_2 \rangle + \langle A_2A_3 \rangle + \langle A_3A_4 \rangle + \langle A_4A_5 \rangle + \langle A_5A'_1 \rangle \geq -3 - \kappa, \quad (20)$$

where  $\kappa$  is a correction term depending on the correlation between  $A_1$  and  $A'_1$ . In our approach, rather than considering these as distinct measurements, we would rather account for any discrepancy as unsharpness and attribute this to either parameter dependence or outcome indeterminism at the hidden variable level. As  $A_1$  is a dichotomic measurement this amounts to attributing  $\eta$  in the following way:

$$\eta = \max_{o \in \{-1,1\}} p(A_1 = o | A'_1 = -o) . \quad (21)$$

The bound of [52] is tighter: they obtain  $\text{CF} \leq 0.041(2)$  versus  $\text{CF} \leq 0.072$  as a criterion for classicality when recast in our terms. A contextuality-by-default analysis led to a normalised bound of  $\text{CF} \leq 0.234$  [27]. In contrast to both analyses we point to the fact that our approach provides a clear relationship with assumptions at the ontological level.

**Ref. [53].** This paper is a photonic implementation of an experiment proposed in [57] which aims at closing the compatibility loophole. It uses photonic qutrit in order to violate the following inequality:

$$P(D_1^A = 1 | D_0^B = 1) - P(T_0^A = a_0 | D_0^B = 1) - P(T_1^A = a_1 | D_0^B = 1) \leq 0, \quad (22)$$

where  $D_i^O$  and  $T_i^O$  are observables and  $a_0$  and  $a_1$  are further parameters that describe the measurements. Here we wish to test our results to compare the experimental results to the theoretical bound. In this paper we can take departure from the expected quantum behaviour as unsharpness at the level of the probabilities, such that:

$$\eta = \max_x (|p_{\text{th}}(x) - p_{\text{exp}}(x)|) . \quad (23)$$

where  $x$  is taken in the set of all possible events:  $D_1^A = 1 | D_0^B = 1$ ,  $T_0^A = a_0 | D_0^B = 1$  and  $T_1^A = a_1 | D_0^B = 1$ .  $p_{\text{th}}$  and  $p_{\text{exp}}$  are the theoretical probability and the experimental probability respectively. The highest difference is found for  $p_{\text{exp}}(T_1^A = a_1 | D_0^B = 1)$ , where the theoretical prediction is 0 and the experimental result is  $0.010 \pm 0.001$ . Then we can assume that  $\eta \leq 0.01$ . On the other hand, signalling is measured to be of the order of  $10^{-3}$  thus we can safely apply Theorem 2 by assuming that  $\sigma = 0.001$ :

$$\text{CF} \leq \eta \leq 0.01, \quad (24)$$

which we can directly compare to their result  $\text{CF} = 0.89 > 0.01$  implying that they are indeed witnessing noise-robust contextuality.

**Ref. [58].** This reference is another significant experiment on contextuality which closes the sharpness, detection and compatibility loopholes, thus claiming to be significantly loophole free. The experiment focuses on an inequality equivalent to the CHSH inequality with two ions of different nature in order to avoid interference on the measurements and operations:

$$\mathcal{C} = \langle \hat{O}_0 \hat{O}_1 \rangle + \langle \hat{O}_1 \hat{O}_2 \rangle + \langle \hat{O}_2 \hat{O}_3 \rangle - \langle \hat{O}_3 \hat{O}_0 \rangle \leq 2. \quad (25)$$

In order to close the sharpness loophole, they measure the repeatability  $R_i$  of the observables  $\hat{O}_i$ . It corresponds to the number of times the observable  $\hat{O}_i$  gives different outcomes when measured twice in a row. The lowest repeatability is found to be  $R_0 \approx 97\%$  for  $\hat{O}_0$ . Taking the repeatability as a good indicator of unsharpness, if we consider that both parties have at least a repeatability  $R_i \geq 1 - \epsilon = 0.97$  then one can define  $\eta = 1 - (1 - \epsilon)^2 = 2\epsilon - \epsilon^2$ . Thus,  $\eta \approx 0.06$ .

We can assume that signalling is so small that it can be neglected. In fact, they measure the cross-talk between the two ion species, and they observe that both have a maximum

population transfer between  $1.9 \times 10^{-6}$  and  $4.3 \times 10^{-6}$  when they are excited by the wrong laser which is negligible for their experiment, and we do not consider it here.

We can compute our bound since  $2\eta + \sigma \approx 0.12 + 0 < 1$  which satisfies the condition for Theorem 2. Then for the empirical model  $e_{\mathcal{C}}$  associated to  $\mathcal{C}$  we have:

$$\text{CF}(e_{\mathcal{C}}) \leq 0.06, \quad (26)$$

which translates in term of the inequality to

$$\begin{aligned} \mathcal{C} &\leq 2 + (4 - 2)\eta \\ &\leq 2.12. \end{aligned} \quad (27)$$

Since they obtain a violation  $\mathcal{C} = 2.526 \pm 0.016$ , our framework with these assumptions certifies this experiment as contextual.

## 7 Conclusion

In this work, we present a new bound on the contextual fraction allowed by hidden-variable models that have access to limited parameter dependence (related to empirical signalling) and nondeterminism (related to empirical unsharpness). We considered several approaches to bounding parameter dependence and nondeterminism of an HVM, which inevitably employ additional assumptions that relate these HV properties to empirical signalling and unsharpness. This allows for a robust testing of non-classicality for noisy measurements.

The main goal of this work is to achieve the above in the broadest and simplest way. It is not our aim to assert which assumptions are valid, though this is a crucial question, but rather to provide clear mathematical results and bounds which are straightforward to check, around which such discussions may take place. A benefit of our approach is that it allows clarity in where the additional assumptions must sit—i.e. in experimentally accessing the values for  $\eta$  and  $\sigma$ . Note that no such assumptions are brought into play yet in proving the bounds of Lemma 2 and Theorem 1; these follow from the definitions. We presented a variety of possible assumptions that would allow  $\eta$  and  $\sigma$  to be empirically bounded by looking at previous works in the literature [38, 47, 52, 53, 54]. In this way, we see that our framework is broad enough to subsume a wide variety of previous approaches.

With respect to applications, and beyond purely foundational considerations, it has also recently been demonstrated that bounds on  $\eta$  and  $\sigma$  can be translated into constraints on adversaries or computational models, for example in generating certified randomness in the presence of bounded signalling on a microchip [17]. Seeking clarity and flexibility in the role and choice of the assumptions that can allow empirical access to parameter dependence and nondeterminism can thus be of benefit for future applications of these inequalities to quantum information and computing more broadly.

## Acknowledgements

The notion of the signalling fraction originally emerged in unpublished work by Samson Abramsky, Rui Soares Barbosa and SM. The authors thank the participants of the 2021 and 2022 editions of the Quantum Contextuality in Quantum Mechanics and Beyond Workshop for their feedback on earlier versions of this work.

This work has received funding from the PEPR integrated project EPiQ ANR-22-PETQ-0007 part of Plan France 2030, and from BPI France Concours Innovation PIA3 projects DOS0148634/00 and DOS0148633/00 – Reconfigurable Optical Quantum Computing.

## Author contibution

K.V. conceptualization, formal analysis (section 4, 5, 6), investigation, visualization, writing – original draft; P-E.E: conceptualization, formal analysis (section 3, 5), investigation, supervision (section 4, 6), visualization, writing – original draft; B.B: validation, writing: review & editing; A.S.: Formal analysis, Validation;S.M. conceptualisation, formal analysis, investigation, supervision, writing: review & editing; D.M. conceptualisation, formal analysis, investigation, supervision, writing: review & editing; S.M. and D.M. contributed equally to this work.

## References

- [1] Bell JS. 1964 On the Einstein Podolsky Rosen paradox. *Physics Physique Fizika* **1**, 195–200. Publisher: American Physical Society ([10.1103/PhysicsPhysiqueFizika.1.195](https://doi.org/10.1103/PhysicsPhysiqueFizika.1.195))
- [2] Bell JS. 1966 On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics* **38**, 447–452. ([10.1103/RevModPhys.38.447](https://doi.org/10.1103/RevModPhys.38.447))
- [3] Simon Kochen ES. 1968 The Problem of Hidden Variables in Quantum Mechanics. *Indiana Univ. Math. J.* **17**, 59–87.
- [4] Kochen S, Specker EP. 1975 The problem of hidden variables in quantum mechanics. In *The logico-algebraic approach to quantum mechanics* , pp. 293–328. Springer.
- [5] Abramsky S, Brandenburger A. 2011 The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics* **13**, 113036. ([10.1088/1367-2630/13/11/113036](https://doi.org/10.1088/1367-2630/13/11/113036))
- [6] Cabello A, Severini S, Winter A. 2014 Graph-theoretic approach to quantum correlations. *Physical Review Letters* **112**, 040401. ([10.1103/PhysRevLett.112.040401](https://doi.org/10.1103/PhysRevLett.112.040401))
- [7] Acín A, Fritz T, Leverrier A, Sainz AB. 2015 A combinatorial approach to non-locality and contextuality. *Communications in Mathematical Physics* **334**, 533–628. ([10.1007/s00220-014-2260-1](https://doi.org/10.1007/s00220-014-2260-1))
- [8] Dzhamfarov EN, Kujala JV. 2016 Context–content systems of random variables: The Contextuality-by-Default theory. *Journal of Mathematical Psychology* **74**, 11–33. ([10.1016/j.jmp.2016.04.010](https://doi.org/10.1016/j.jmp.2016.04.010))
- [9] Raussendorf R. 2013 Contextuality in measurement-based quantum computation. *Physical Review A* **88**, 022322. ([10.1103/PhysRevA.88.022322](https://doi.org/10.1103/PhysRevA.88.022322))
- [10] Howard M, Wallman J, Veitch V, Emerson J. 2014 Contextuality supplies the ‘magic’ for quantum computation. *Nature* **510**, 351–355. ([10.1038/nature13460](https://doi.org/10.1038/nature13460))
- [11] Abramsky S, Barbosa RS, Mansfield S. 2017 Contextual fraction as a measure of contextuality. *Physical Review Letters* **119**, 050504. ([10.1103/PhysRevLett.119.050504](https://doi.org/10.1103/PhysRevLett.119.050504))
- [12] Bermejo-Vega J, Delfosse N, Browne DE, Okay C, Raussendorf R. 2017 Contextuality as a resource for models of quantum computation with qubits. *Physical Review Letters* **119**, 120505. ([10.1103/PhysRevLett.119.120505](https://doi.org/10.1103/PhysRevLett.119.120505))
- [13] Abramsky S, Barbosa RS, de Silva N, Zapata O. 2017 The quantum monad on relational structures. In Larsen KG, Bodlaender HL, Raskin JF, editors, *42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017)* vol. 83 *Leibniz International Proceedings in Informatics (LIPIcs)* pp. 35:1–35:19. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ([10.4230/LIPIcs.MFCS.2017.35](https://doi.org/10.4230/LIPIcs.MFCS.2017.35))

- [14] Mansfield S, Kashefi E. 2018 Quantum advantage from sequential-transformation contextuality. *Physical Review Letters* **121**, 230401. ([10.1103/PhysRevLett.121.230401](https://arxiv.org/abs/10.1103/PhysRevLett.121.230401))
- [15] Karanjai A, Wallman JJ, Bartlett SD. 2018 Contextuality bounds the efficiency of classical simulation of quantum processes. Preprint [arXiv:1802.07744](https://arxiv.org/abs/1802.07744) [quant-ph].
- [16] Bowles J, Wright VJ, Farkas M, Killoran N, Schuld M. 2023 Contextuality and inductive bias in quantum machine learning. *arXiv preprint arXiv:2302.01365*.
- [17] Fyrillas A, Bourdoncle B, Mainos A, Emeriau PE, Start K, Margaria N, Morassi M, Lemaître A, Sagnes I, Stepanov P et al.. 2023 Certified randomness in tight space. *arXiv preprint arXiv:2301.03536*.
- [18] Barrett J, Kent A, Pironio S. 2006 Maximally Nonlocal and Monogamous Quantum Correlations. *Physical Review Letters* **97**, 170409. ([10.1103/PhysRevLett.97.170409](https://arxiv.org/abs/10.1103/PhysRevLett.97.170409))
- [19] Amsellem E, Danielsen LE, López-Tarrida AJ, Portillo JR, Bourennane M, Cabello A. 2012 Experimental Fully Contextual Correlations. *Physical Review Letters* **108**, 200405. ([10.1103/PhysRevLett.108.200405](https://arxiv.org/abs/10.1103/PhysRevLett.108.200405))
- [20] Spekkens RW. 2014 The status of determinism in proofs of the impossibility of a noncontextual model of quantum theory. *Foundations of Physics* **44**, 1125–1155. ([10.1007/s10701-014-9833-x](https://arxiv.org/abs/10.1007/s10701-014-9833-x))
- [21] Leifer M. 2014 Is the Quantum State Real? An Extended Review of  $\psi$ -ontology Theorems. *Quanta* **3**, 67–155. ([10.12743/quanta.v3i1.22](https://arxiv.org/abs/10.12743/quanta.v3i1.22))
- [22] Chiribella G, Yuan X. 2014 Measurement sharpness cuts nonlocality and contextuality in every physical theory. Preprint [arXiv:1404.3348](https://arxiv.org/abs/1404.3348) [quant-ph]. ([10.48550/arXiv.1404.3348](https://arxiv.org/abs/10.48550/arXiv.1404.3348))
- [23] Zhang Y, Glancy S, Knill E. 2011 Asymptotically Optimal Data Analysis for Rejecting Local Realism. *Physical Review A* **84**, 062118. ([10.1103/PhysRevA.84.062118](https://arxiv.org/abs/10.1103/PhysRevA.84.062118))
- [24] Zhang Y, Glancy S, Knill E. 2013 Efficient Quantification of Experimental Evidence against Local Realism. *Physical Review A* **88**, 052119. ([10.1103/PhysRevA.88.052119](https://arxiv.org/abs/10.1103/PhysRevA.88.052119))
- [25] Dupuis F, Fawzi O, Renner R. 2020 Entropy Accumulation. *Communications in Mathematical Physics* **379**, 867–913. ([10.1007/s00220-020-03839-5](https://arxiv.org/abs/10.1007/s00220-020-03839-5))
- [26] Lin PS, Rosset D, Zhang Y, Bancal JD, Liang YC. 2018 Device-Independent Point Estimation from Finite Data and Its Application to Device-Independent Property Estimation. *Physical Review A* **97**, 032309. ([10.1103/PhysRevA.97.032309](https://arxiv.org/abs/10.1103/PhysRevA.97.032309))
- [27] Kujala JV, Dzhafarov EN, Larsson JÅ. 2015 Necessary and sufficient conditions for an extended noncontextuality in a broad class of quantum mechanical systems. *Physical Review Letters* **115**, 150401.
- [28] Wang D, Sadrzadeh M, Abramsky S, Víctor H, Cervantes V. 2021a Analysing ambiguous nouns and verbs with quantum contextuality tools. *Journal of Cognitive Science* **22**, 391–420.
- [29] Wang D, Sadrzadeh M, Abramsky S, Cervantes VH. 2021b On the quantum-like contextuality of ambiguous phrases. *arXiv preprint arXiv:2107.14589*.
- [30] Cervantes VH, Dzhafarov EN. 2018 Snow queen is evil and beautiful: Experimental evidence for probabilistic contextuality in human choices.. *Decision* **5**, 193.
- [31] Basieva I, Cervantes VH, Dzhafarov EN, Khrennikov A. 2019 True contextuality beats direct influences in human decision making.. *Journal of Experimental Psychology: General* **148**, 1925.

- [32] Cervantes VH. 2023 A note on the relation between the Contextual Fraction and CNT2. *Journal of Mathematical Psychology* **112**, 102726.
- [33] Camillo G, Cervantes VH. 2023 Measures of contextuality in cyclic systems and the negative probabilities measure CNT3. *arXiv preprint arXiv:2305.16574*.
- [34] Kujala JV, Dzhafarov EN. 2019 Measures of contextuality and non-contextuality. *Philosophical Transactions of the Royal Society A* **377**, 20190149.
- [35] Kunjwal R, Spekkens RW. 2015 From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism. *Physical Review Letters* **115**, 110403. ([10.1103/PhysRevLett.115.110403](https://doi.org/10.1103/PhysRevLett.115.110403))
- [36] Kunjwal R. 2019 Beyond the Cabello-Severini-Winter framework: Making sense of contextuality without sharpness of measurements. *Quantum* **3**, 184. ([10.22331/q-2019-09-09-184](https://doi.org/10.22331/q-2019-09-09-184))
- [37] Kunjwal R, Spekkens RW. 2018 From statistical proofs of the Kochen-Specker theorem to noise-robust noncontextuality inequalities. *Physical Review A* **97**, 052110. ([10.1103/PhysRevA.97.052110](https://doi.org/10.1103/PhysRevA.97.052110))
- [38] Winter A. 2014 What does an experimental test of quantum contextuality prove or disprove?. *Journal of Physics A: Mathematical and Theoretical* **47**, 424031. ([10.1088/1751-8113/47/42/424031](https://doi.org/10.1088/1751-8113/47/42/424031))
- [39] Kirchmair G, Zähringer F, Gerritsma R, Kleinmann M, Gühne O, Cabello A, Blatt R, Roos CF. 2009 State-independent experimental test of quantum contextuality. *Nature* **460**, 494–497. ([10.1038/nature08172](https://doi.org/10.1038/nature08172))
- [40] Kleinmann M, Gühne O, Portillo JR, Larsson JÅ, Cabello A. 2011 Memory cost of quantum contextuality. *New Journal of Physics* **13**, 113011. ([10.1088/1367-2630/13/11/113011](https://doi.org/10.1088/1367-2630/13/11/113011))
- [41] Hall MJW. 2011 Relaxed Bell Inequalities and Kochen-Specker Theorems. *Physical Review A* **84**, 022102. ([10.1103/PhysRevA.84.022102](https://doi.org/10.1103/PhysRevA.84.022102))
- [42] Spekkens RW. 2005 Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A* **71**, 052108. ([10.1103/PhysRevA.71.052108](https://doi.org/10.1103/PhysRevA.71.052108))
- [43] Popescu S, Rohrlich D. 1994 Quantum nonlocality as an axiom. *Foundations of Physics* **24**, 379–385. ([10.1007/bf02058098](https://doi.org/10.1007/bf02058098))
- [44] Norsen T. 2006 Bell Locality and the Nonlocal Character of Nature. *Foundations of Physics Letters* **19**, 633–655. ([10.1007/s10702-006-1055-9](https://doi.org/10.1007/s10702-006-1055-9))
- [45] Boyd S, Boyd SP, Vandenberghe L. 2004 *Convex optimization*. Cambridge university press. ([10.1017/cbo9780511804441](https://doi.org/10.1017/cbo9780511804441))
- [46] Araújo M, Quintino MT, Budroni C, Cunha MT, Cabello A. 2013 All noncontextuality inequalities for the  $n$ -cycle scenario. *Physical Review A* **88**, 022118. ([10.1103/PhysRevA.88.022118](https://doi.org/10.1103/PhysRevA.88.022118))
- [47] Gühne O, Kleinmann M, Cabello A, Larsson JÅ, Kirchmair G, Zähringer F, Gerritsma R, Roos CF. 2010 Compatibility and noncontextuality for sequential measurements. *Physical Review A* **81**, 022121. ([10.1103/PhysRevA.81.022121](https://doi.org/10.1103/PhysRevA.81.022121))
- [48] Fine A. 1982 Hidden Variables, Joint Probability, and the Bell Inequalities. *Physical Review Letters* **48**, 291–295. ([10.1103/PhysRevLett.48.291](https://doi.org/10.1103/PhysRevLett.48.291))

- [49] Clemente L, Kofler J. 2015 Necessary and sufficient conditions for macroscopic realism from quantum mechanics. *Physical Review A* **91**, 062103. ([10.1103/PhysRevA.91.062103](#))
- [50] Moreira SV, Cunha MT. 2019 Quantifying quantum invasiveness. *Physical Review A* **99**, 022124. ([10.1103/PhysRevA.99.022124](#))
- [51] Spekkens RW, Buzacott DH, Keehn AJ, Toner B, Pryde GJ. 2009 Preparation contextuality powers parity-oblivious multiplexing. *Physical Review Letters* **102**, 010401. ([10.1103/PhysRevLett.102.010401](#))
- [52] Lapkiewicz R, Li P, Schaeff C, Langford NK, Ramelow S, Wiesniak M, Zeilinger A. 2011 Experimental non-classicality of an indivisible quantum system. *Nature* **474**, 490–493. ([10.1038/nature10119](#))
- [53] Hu XM, Chen JS, Liu BH, Guo Y, Huang YF, Zhou ZQ, Han YJ, Li CF, Guo GC. 2016 Experimental Test of Compatibility-Loophole-Free Contextuality with Spatially Separated Entangled Qutrits. *Physical Review Letters* **117**, 170403. ([10.1103/PhysRevLett.117.170403](#))
- [54] Marques B, Ahrens J, Nawareg M, Cabello A, Bourennane M. 2014 Experimental Observation of Hardy-Like Quantum Contextuality. *Physical Review Letters* **113**, 250403. ([10.1103/PhysRevLett.113.250403](#))
- [55] Mermin ND. 1993 Hidden variables and the two theorems of John Bell. *Reviews of Modern Physics* **65**, 803–815. ([10.1103/RevModPhys.65.803](#))
- [56] Klyachko AA, Can MA, Binicioğlu S, Shumovsky AS. 2008 Simple Test for Hidden Variables in Spin-1 Systems. *Physical Review Letters* **101**, 020403. ([10.1103/PhysRevLett.101.020403](#))
- [57] Cabello A, Cunha MT. 2011 Proposal of a Two-Qutrit Contextuality Test Free of the Finite Precision and Compatibility Loopholes. *Physical Review Letters* **106**, 190401. ([10.1103/PhysRevLett.106.190401](#))
- [58] Wang P, Zhang J, Luan CY, Um M, Wang Y, Qiao M, Xie T, Zhang JN, Cabello A, Kim K. 2022 Significant Loophole-Free Test of Kochen-Specker Contextuality Using Two Species of Atomic Ions. *Science Advances* **8**, eabk1660. ([10.1126/sciadv.abk1660](#))
- [59] Barbosa RS. 2014 On monogamy of non-locality and macroscopic averages: examples and preliminary results. In Coecke B, Hasuo I, Panangaden P, editors, *11th Workshop on Quantum Physics and Logic (QPL 2014)* vol. 172 *Electronic Proceedings in Theoretical Computer Science* pp. 36–55. Open Publishing Association. ([10.4204/EPTCS.172.4](#))
- [60] Barbosa RS. 2015 *Contextuality in quantum mechanics and beyond*. PhD thesis University of Oxford.
- [61] Carù G. 2018 Towards a complete cohomology invariant for non-locality and contextuality. Preprint [arXiv:1807.04203 \[quant-ph\]](#).
- [62] Karvonen M. 2019 Categories of Empirical Models. In Selinger P, Chiribella G, editors, *15th International Conference on Quantum Physics and Logic (QPL 2018)* vol. 287 *Electronic Proceedings in Theoretical Computer Science* pp. 239–252. Open Publishing Association. ([10.4204/EPTCS.287.14](#))
- [63] Abramsky S, Barbosa RS, Karvonen M, Mansfield S. 2019 A comonadic view of simulation and quantum resources. To appear in Proceedings of the 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS 2019) ([10.1109/lics.2019.8785677](#))

- [64] Clauser JF, Horne MA, Shimony A, Holt RA. 1969 Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters* **23**, 880–884. ([10.1103/PhysRevLett.23.880](#))
- [65] Barbosa RS, Douce T, Emeriau PE, Kashefi E, Mansfield S. 2019 Continuous-variable nonlocality and contextuality. Preprint [arXiv:1905.082671 \[quant-ph\]](#).
- [66] Silman J, Pironio S, Massar S. 2013 Device-Independent Randomness Generation in the Presence of Weak Cross-Talk. *Physical Review Letters* **110**, 100504.
- [67] Barvinok A. 2002 *A course in convexity* vol. 54 *Graduate Studies in Mathematics*. American Mathematical Society. ([10.1090/gsm/054](#))
- [68] Clauser JF, Horne MA, Shimony A, Holt RA. 1969 Proposed Experiment to Test Local Hidden-Variable Theories. *Physical Review Letters* **23**, 880–884. ([10.1103/PhysRevLett.23.880](#))

# Supplemental Material

## A General framework for contextuality

This section summarises some of the main ideas from [5, 11], setting out the definition of contextuality, assumptions on hidden variable models, and how these are related, and finally defining the *contextual fraction*, a quantitative measure of the degree to which observed empirical data is contextual.

### A.1 Scenarios and Behaviours

An abstract description of a particular experimental setup is formalised as a *measurement scenario*.

**Definition 3** (Measurement scenario). *A measurement scenario is a triple  $\langle X, \mathcal{M}, O \rangle$  where:*

- $X$  is a finite set of measurement labels.
- $\mathcal{M}$  is a covering family of  $X$  i.e. it is a set of subsets of  $X$  such that  $\bigcup_{C \in \mathcal{M}} C = X$ . The element  $C \in \mathcal{M}$  are taken as maximal contexts and represent maximal sets of compatible observables<sup>7</sup>.
- $O = (O_x)_{x \in X}$  is a finite set of outcomes for each measurement. If some set of measurements  $U \subseteq X$  are considered together then the corresponding joint outcome set is given by the (Cartesian) product of the respective outcome spaces:  $O_U = \prod_{x \in U} O_x$ .

For example, the setup corresponding to the usual CHSH experiment [64] or (2,2,2) Bell scenario—two spacelike separated parties, Alice and Bob, each have two possible binary measurements—is described by:

$$\begin{aligned} X &= \{a, a', b, b'\}, \\ \mathcal{M} &= \{ \{a, b\}, \{a, b'\}, \{a', b\}, \{a', b'\} \}, \\ \forall x \in X, O_x &= \{0, 1\}. \end{aligned} \tag{A.28}$$

Next, given the description of the experimental setup, either calculating theoretical predictions for admissible joint outcomes or performing repeated runs of the experiment with varying choices of measurement context and recording the frequencies of the corresponding joint events results in a probability table which is formalised as an *empirical model*.

**Definition 4** (Empirical model/behaviour). *Given a measurement scenario  $\langle X, \mathcal{M}, O \rangle$ , an empirical model (or behaviour) is a family  $e = (e_C)_{C \in \mathcal{M}}$  where  $e_C$  is a probability distribution on the joint outcome space  $O_C$ .*

For example, an empirical model corresponding to the usual CHSH experiment is given by Table 2. We may further require that the empirical data is non-signalling by imposing compatibility conditions across different contexts, that is, for  $C, C' \in \mathcal{M}$ ,  $e_C|_{C \cap C'} = e_{C'}|_{C \cap C'}$ .

---

<sup>7</sup>From recent works using the Sheaf-theoretic treatments for contextuality [59, 60, 61, 62, 63],  $\mathcal{M}$  is often taken as the set of maximal contexts requiring that  $\mathcal{M}$  be an anti-chain with respect to subset inclusion i.e. that no element of the family is a proper subset of another.

A	B	00	01	10	11
$a$	$b$	$p_1$	$p_2$	$p_2$	$p_1$
$a$	$b'$	$p_1$	$p_2$	$p_2$	$p_1$
$a'$	$b$	$p_1$	$p_2$	$p_2$	$p_1$
$a'$	$b'$	$p_2$	$p_1$	$p_1$	$p_2$

Table 2: Quantum empirical model on the  $(2, 2, 2)$  Bell scenario specifying the probabilities of the joint outcomes for the CHSH model with  $p_1 = \frac{2+\sqrt{2}}{8}$  and  $p_2 = \frac{2-\sqrt{2}}{8}$ . Here  $a_i$  and  $b_i$  for  $i = 1, 2$  represent quantum observables for CHSH with eigenvalues relabelled as 0 and 1. Joint probabilities are obtained by the Born rule.

## A.2 Contextuality and its quantification via the contextual fraction

Now that we have defined what we mean by a measurement scenario and an empirical model we can define precisely what we mean by *contextuality* which is a property of the empirical data.

**Definition 5** (Noncontextuality). *A non-signalling empirical model  $e$  on a measurement scenario  $\langle X, \mathcal{M}, O \rangle$  is said to be noncontextual if there exists a global probability distribution  $d$  on  $O_X$  such that  $e_C = d|_C$ .*

This corresponds to a property of extendability for the empirical data i.e. the fact that we can extend all probability distributions into a global one from which it is possible to marginalise in order to retrieve a probability distribution for a given context. If such a global distribution cannot be found then  $e$  is said to be *contextual*.

A more refined question than asking if an empirical model  $e$  is contextual is to ask what fraction of the empirical data admits a noncontextual explanation. This is formalised as the *contextual fraction* [5, 11]. Instead of asking for a global probability distribution from which the empirical behaviour  $e$  can be deduced by marginalisation at each context, the idea is to look for a subprobability distribution  $b$  (i.e. that sums to less than 1) on  $O_X$  explaining some fraction of the data i.e. we require that  $\forall C \in \mathcal{M}, b|_C \leq e_C$ . Ideally, we would like to find such a  $b$  maximising  $\mathbf{1} \cdot b = \sum_{i=1}^{|O_X|} b_i$ .

**Definition 6** (Noncontextual fraction). *Given an empirical model  $e$  on a measurement scenario  $\langle X, \mathcal{M}, O \rangle$ , its noncontextual fraction is given by:*

$$\text{NCF}(e) := \sup \{ \mathbf{1} \cdot b \mid b \geq 0 \text{ s.t. } \forall C \in \mathcal{M}, b|_C \leq e_C \} .$$

Note that by construction, we have that  $\text{NCF}(e) \in [0, 1]$ . The case  $\text{NCF}(e) = 1$  for non-signalling data corresponds to noncontextuality while  $\text{NCF}(e) = 0$  corresponds to maximal contextuality or strong contextuality [5]. Naturally, the contextual fraction is defined as:  $\text{CF}(e) := 1 - \text{NCF}(e)$ . We would like to stress that for non-signalling data contextuality corresponds to the case that  $\text{CF}(e) > 0$ , though our aim is to modify this inequality to include about genuine contextuality of noisy data.

Now we would like to give a more geometric take on the contextual fraction. Given two empirical models  $e_1$  and  $e_2$  on the same measurement scenario  $\langle X, \mathcal{M}, O \rangle$ , we can define a third valid empirical by taking a convex sum of  $e_1$  and  $e_2$ : for instance  $\lambda e_1 + (1 - \lambda)e_2$  for  $\lambda \in [0, 1]$  is another valid empirical model. Asking what fraction of a non-signalling empirical model  $e$  admits a noncontextual explanation is asking for a convex decomposition of the form:

$$e = \lambda e^{\text{NC}} + (1 - \lambda)e' \tag{A.29}$$

with weight  $\lambda$  on the noncontextual part. Maximising  $\lambda$  yields the noncontextual fraction  $\text{NCF}(e)$  (see Figure 3).

A convenient property of the contextual fraction is that its computation can be cast as a linear program [11]. Fix  $n := |O_X|$  the number of different global assignments and  $m := \sum_{C \in \mathcal{M}} |O_C|$  the number of total local assignments ranging over contexts. We use bold notation for vectors. Local assignments can be listed as:  $\{\langle C, \mathbf{s} \rangle \text{ s.t. } C \in \mathcal{M} \text{ and } \mathbf{s} \in O_C\}$ . Then the incidence matrix  $M$  which records the possible restrictions from global to local assignments is a  $m \times n$  (0,1)-matrix defined as:

$$M[\langle C, \mathbf{s} \rangle, \mathbf{g}] := \begin{cases} 1 & \text{if } \mathbf{g}|_C = \mathbf{s} \\ 0 & \text{otherwise} . \end{cases} \quad (\text{A.30})$$

To understand its action, read  $M$  column-by-column by fixing a global section  $\mathbf{g} \in O_X$ . For this specific column,  $M$  assigns the value 1 in the row corresponding to the local section  $\mathbf{s}$  in context  $C$  whenever  $\mathbf{g}|_C = \mathbf{s}$  (and 0 otherwise). The empirical data  $e$  can be represented as a vector  $\mathbf{v}^e \in [0, 1]^m$  where for a given context  $C \in \mathcal{M}$  and a local assignment  $\mathbf{s} \in O_C$ ,  $\mathbf{v}^e[\langle C, \mathbf{s} \rangle] = e_C(\mathbf{s})$ . This is the flattened version of the tables (see for instance Table 2) that are usually used to represent empirical models. In the noise-free regime  $e$  is noncontextual if and only if there exists a global probability distribution  $\mathbf{d} \in [0, 1]^n$  such that  $M\mathbf{d} = \mathbf{v}^e$  and that  $\text{NCF}(e) < 1$  corresponds to the non-existence of such a global distribution  $\mathbf{d}$ . Definition 6 can be expanded as:

$$(\text{P-NCF}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{b} \in \mathbb{R}^n \\ \text{maximising } \mathbf{1} \cdot \mathbf{b} \\ \text{subject to:} \\ \quad M\mathbf{b} \leq \mathbf{v}^e \\ \quad \mathbf{b} \geq \mathbf{0} . \end{array} \right.$$

Interestingly, the dual of this program (with a clever change of variable) is of particular interest since it allows to compute the expression of a Bell inequality optimised to the empirical data [11, 65]. We present the dual program of (P-NCF) without the change of variable  $a := |\mathcal{M}|^{-1}\mathbf{1} - \mathbf{y}$  below as it is enough to prove the continuity result of interest.

$$(\text{D-NCF}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{y} \in \mathbb{R}^m \\ \text{minimising } \mathbf{y} \cdot \mathbf{v}^e \\ \text{subject to:} \\ \quad M^T \mathbf{y} \geq \mathbf{1} \\ \quad \mathbf{y} \geq \mathbf{0} . \end{array} \right.$$

### A.3 Quantifying (non-)signalling

For empirical models outside of the non-signalling polytope, it is possible to define a measure to know *how much the data is signalling* or *how far away from the non-signalling polytope* the empirical model is. This measure has similar properties to the contextual fraction but instead of only using global value assignments which are given by the vertices of the noncontextual polytope (see Figure 3), we also admit maximally contextual vertices from the non-signalling polytope (e.g. PR-boxes in the (2, 2, 2) Bell scenario). To reach those vertices, we allow the global distribution to have negative weight on deterministic noncontextual vertices as long as the marginals for each context are nonnegative. It follows from [5, Theorems 5.5 and 5.9]. We call this measure the *non-signalling fraction* NSF.

**Definition 7** (non-signalling fraction). *Given an empirical model  $e$  on a measurement scenario  $\langle X, \mathcal{M}, O \rangle$ , its non-signalling fraction is given by:*

$$\text{NSF}(e) := \sup \{ \mathbf{1} \cdot \mathbf{b} \mid \forall C \in \mathcal{M}, 0 \leq b|_C \leq e_C \} .$$

The signalling fraction of  $e$  is the given by  $\text{SF}(e) := 1 - \text{NSF}(e)$ . NSF represents the irreducible weight on signalling vertices one must have in order to explain the data. Similar to the contextual fraction, it bears a geometrical interpretation (see Figure 3) Asking what fraction of the empirical model  $e$  admits a non-signalling explanation is asking for a convex decomposition of the form:

$$e = \lambda e^{\text{NS}} + (1 - \lambda) e', \quad (\text{A.31})$$

with weight  $\lambda$  on the non-signalling part. Maximising  $\lambda$  yields the non-signalling fraction  $\text{NSF}(e)$ .

Idealised models predicted by quantum mechanics, where contexts consist of perfectly commuting observables, should not be signalling; however, in realistic settings noise can manifest itself to make empirical models somewhat signalling. The fraction defined above will capture this effect as a property of the empirical data. The constraint that for all contexts  $C$ ,  $b|_C$  must be nonnegative ensures that, while probabilities might be negative, they sum to give a valid empirical model. Just like the contextual fraction, the signalling fraction of any empirical behaviour can be efficiently computed through a linear program:

$$(\text{P-NSF}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{b} \in \mathbb{R}^n \\ \text{maximising } \mathbf{1} \cdot \mathbf{b} \\ \text{subject to:} \\ \quad \mathbf{M}\mathbf{b} \leq \mathbf{v}^e \\ \quad \mathbf{M}\mathbf{b} \geq 0 . \end{array} \right.$$

Signalling is often associated to disagreeing marginals [47, 66] hence the following quantification:

**Definition 8** (Maximum Incompatibility of Marginals (MIM)). *The MIM of an empirical model  $e$  in a measurement scenario  $\langle X, \mathcal{M}, O \rangle$  is given by the highest incompatibility of marginals:*

$$\text{MIM}(e) = \max \left\{ \left| e_C|_{C \cap C'}(t) - e_{C'}|_{C \cap C'}(t) \right|, \forall C \neq C' \in \mathcal{M}, \forall t \in O_{C \cap C'} \right\}, \quad (\text{A.32})$$

where  $O_{C \cap C'}$  is the outcome of a measurement in the intersection of  $C$  and  $C'$  (see A.1).

Both notions of signalling (MIM and SF) agree on the non-signalling polytope (i.e. they are both equal to 0). However there exist points  $e$  that are maximally signalling in the sense that they belong to a facet of the signalling polytope thus  $\text{SF}(e) = 1$  and for which  $\text{MIM}(e) < 1$ . For instance the following empirical model has  $\text{SF} = 1$  but  $\text{MIM} = 0.2821$ :

A	B	00	01	10	11
$a_1$	$b_1$	0	0	0	1
$a_1$	$b_2$	0.2821	0	0.0674	0.6505
$a_2$	$b_1$	0.2821	0.0674	0	0.6505
$a_2$	$b_2$	0.0821	0.4589	0.4589	0

Given its geometric interpretation, we believe the signalling-fraction is a more natural quantification of signalling. We prove below that MIM is always a lower-bound for SF.

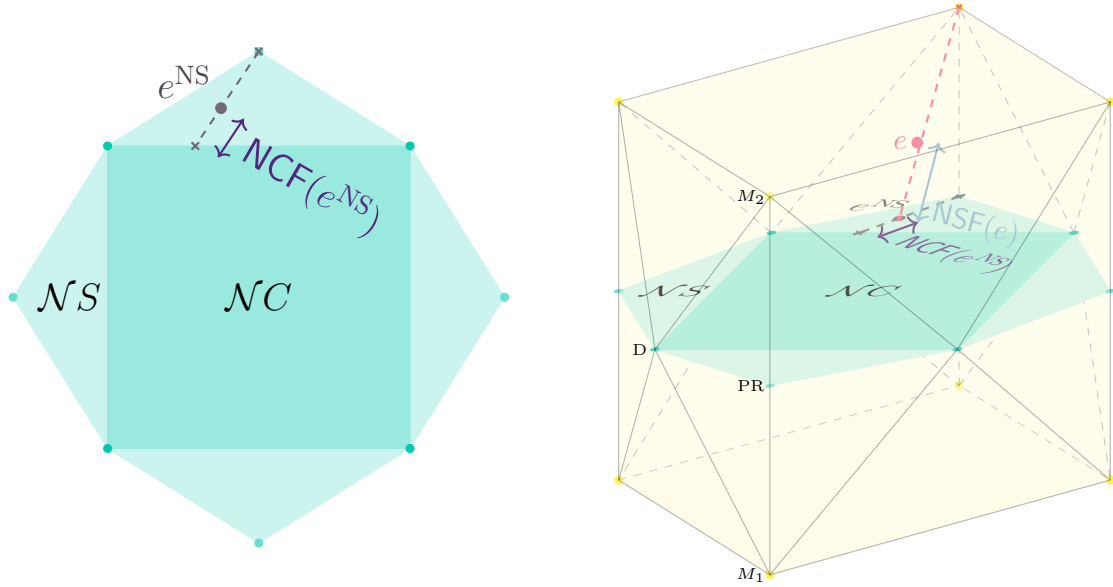


Figure 3: [Left] A schematic representation of a 2-dimensional cut of the non-signalling correlation polytope for the (2,2,2) Bell scenario. Inscribed in it, in darker colour, is the local (or noncontextual) polytope of correlations. Vertices of the polytopes are highlighted. The quantum boundary (not represented) would correspond to a convex line between the two polytopes. The noncontextual fraction of  $e$   $NCF(e)$  is represented and we have  $CF(e) := 1 - NCF(e)$ . [Right] A schematic representation of a cut of the signalling correlation polytope for the (2,2,2) Bell scenario with the non-signalling (NS) polytope and the noncontextual (NC) polytope inscribed in it. Vertices of the polytopes are highlighted. Note that contextual vertices of the non-signalling polytope, unlike noncontextual ones, are not vertices of the signalling polytope. For a signalling behaviour  $e$ , its non-signalling fraction (maximal weight on the non-signalling part  $e^{\text{NS}}$  in a convex decomposition) is represented. Likewise, the noncontextual fraction of  $e^{\text{NS}}$  is represented.

**Lemma 2.** *The signalling fraction of a behaviour  $e$  is lower bounded by its MIM.*

*Proof.* For any behaviour  $e$ , by definition of the non-signalling fraction, decompose it as follows:

$$e = \text{NSF}(e)e^{\text{NS}} + \text{SF}(e)e', \quad (\text{A.33})$$

with  $e^{\text{NS}}$  a non-signalling model. By convexity of the MIM measure:

$$\text{MIM}(e) \leq \text{NSF}(e)\text{MIM}(e^{\text{NS}}) + \text{SF}(e)\text{MIM}(e') \quad (\text{A.34})$$

$$\leq \text{SF}(e)\text{MIM}(e') \quad (\text{A.35})$$

$$\leq \text{SF}(e), \quad (\text{A.36})$$

since  $\text{MIM}(e^{\text{NS}}) = 0$  and  $\text{MIM}(e') \leq 1$ .  $\square$

#### A.4 Hidden-variable models

Hidden variable models (HVMs) provide a broad general approach to considering possible physical explanations underlying observed data, often with the goal of explaining certain properties in a more intuitive way; e.g. analogous to how probabilistic behaviours of classical statistical mechanics admit a deeper, albeit more complex, deterministic description. Thus the assumptions that we place on HVMs are usually ways of encoding notions of classicality.

**Definition 9** (Hidden-variable model). *A hidden-variable model on a measurement scenario  $\langle X, \mathcal{M}, O \rangle$  consists of the triple  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  where:*

- $\Lambda$  is the finite space of hidden variables.
- $p$  is a probability distribution on  $\Lambda$ .
- for each  $\lambda \in \Lambda$ ,  $h^\lambda$  is empirical model i.e.  $h^\lambda = (h_C^\lambda)_{C \in \mathcal{M}}$  is a family where, for each context  $C \in \mathcal{M}$ ,  $h_C^\lambda$  is a probability distribution on the joint outcome space  $\mathcal{O}_C$ .

This gives rise to a behaviour  $h := \sum_{\lambda \in \Lambda} p(\lambda) h^\lambda$ .

We say that  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  is *parameter-independent* whenever  $\forall \lambda$  and for any two contexts  $C_1$  and  $C_2$ , we have that  $h_{C_1}^\lambda|_{C_1 \cap C_2} = h_{C_2}^\lambda|_{C_1 \cap C_2}$ . This ensures that the behaviour  $h$  is non-signalling.

We say that  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  is *outcome deterministic* whenever  $\forall \lambda$ ,  $h^\lambda$  gives a deterministic joint outcome for each context (in other words, it is a vertex of the polytope).

The following proposition is a corollary of [5, Proposition 3.1 and Theorem 8.1].

**Proposition 2.** *An empirical model  $e$  is noncontextual ( $\text{CF}(e) = 0$ ) if and only if it is realisable by an outcome deterministic parameter-independent HVM.*

The set of local behaviour  $\mathcal{L}$  is described by a convex combination of hidden variable behaviours which are deterministic and non-signalling. In other words:

**Definition 10** (local polytope). *The set of local behaviour  $\mathcal{L}$  is defined as:*

$$\mathcal{L} := \left\{ h = \sum_{\lambda \in \Lambda} p(\lambda) h^\lambda \mid \langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle \text{ is PI and OD} \right\}. \quad (\text{A.37})$$

This exactly corresponds to the set of behaviours for which the contextual fraction is exactly zero. We start by relaxing determinism. We call  $\eta$  the fraction of outcome nondeterminism.

**Definition 11** ( $(1-\eta)$  OD HVM). *A parameter-independent hidden-variable model is said to be  $(1-\eta)$  outcome deterministic if for all hidden variables  $\lambda$ , there exists a decomposition of  $h^\lambda$  of the form  $h^\lambda = (1-\eta_\lambda)h_{\text{OD}}^\lambda + \eta_\lambda h''^\lambda$  with  $h_{\text{OD}}^\lambda$  an outcome deterministic model, potentially signalling, and with necessarily  $\eta_\lambda \leq \eta$ .*

This definition ensures that a PR-box [43] is unsharp since it only admits a  $(1-\eta)$  outcome deterministic HVM with  $\eta = 0.5$  when the hidden variables are parameter-independent. Taking back the measurement scenario for CHSH (A.28), a PR-box has a probability table given by Table 3.

We now relax parameter-independence:

**Definition 12** ( $(1-\sigma)$  PI HVM). *A hidden-variable model is said to be  $(1-\sigma)$  parameter-independent if for all hidden variables  $\lambda$ , there exists a decomposition of  $h^\lambda$  of the form  $h^\lambda = (1-\sigma_\lambda)h_{\text{NS}}^\lambda + \sigma_\lambda h'^\lambda$  with  $h_{\text{NS}}^\lambda$  a parameter-independent model and with necessarily  $\sigma_\lambda \leq \sigma$ .*

If one releases parameter independence then there are many possible decompositions of the PR-box given by Table 3. We give one example of such HVM decomposition with parameter-dependent hidden variable behaviour in Table 4.



$\max\{0, -\omega(o)\}$ . Then Eq. (B.40) becomes:

$$\forall C \in \mathcal{M}, \sum_{o \in O_C} \omega_C^+(o) = \sum_{o \in O_C} \omega_C^-(o) . \quad (\text{B.42})$$

And, because  $\omega^+$  and  $\omega^-$  have disjoint support, Eq. (B.41) becomes:

$$\forall C \in \mathcal{M}, \sum_{o \in O_C} \omega_C^+(o) + \sum_{o \in O_C} \omega_C^-(o) \leq 2\varepsilon . \quad (\text{B.43})$$

Eq. (B.42) and (B.43) imply that:

$$\forall C \in \mathcal{M}, \sum_{o \in O_C} \omega_C^\pm(o) \leq \varepsilon . \quad (\text{B.44})$$

We also define  $\mathbf{v}^\omega$  (resp.  $\mathbf{v}^{\omega^+}$  and  $\mathbf{v}^{\omega^-}$ ) to be the flattened version of the table  $\omega$  (resp.  $\omega^+$  and  $\omega^-$ ) as defined in the previous section. From, Eq. (B.44):

$$\|\mathbf{v}^{\omega^\pm}\|_1 = \mathbf{1} \cdot \mathbf{v}^{\omega^\pm} = \sum_{C \in \mathcal{M}} \sum_{o \in O_C} \omega^\pm(o) \leq |\mathcal{M}|\varepsilon . \quad (\text{B.45})$$

The noncontextual fraction of  $e$   $\text{NCF}(e)$  is the optimal value of the linear program:

$$(\text{P-CF}^e) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{b} \in \mathbb{R}^n \\ \text{maximising } \mathbf{1} \cdot \mathbf{b} \\ \text{subject to:} \\ \quad \mathbf{M}\mathbf{b} \leq \mathbf{v}^e \\ \quad \mathbf{b} \geq \mathbf{0} . \end{array} \right.$$

Its dual program reads:

$$(\text{D-CF}^e) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{y} \in \mathbb{R}^m \\ \text{minimising } \mathbf{y} \cdot \mathbf{v}^e \\ \text{subject to:} \\ \quad \mathbf{M}^T \mathbf{y} \geq \mathbf{1} \\ \quad \mathbf{y} \geq \mathbf{0} . \end{array} \right.$$

To compute the noncontextual fraction of  $e'$   $\text{NCF}(e')$ , it suffices to replace  $\mathbf{v}^e$  by  $\mathbf{v}^{e'}$  in the programs above:

$$(\text{P-CF}^{e'}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{b} \in \mathbb{R}^n \\ \text{maximising } \mathbf{1} \cdot \mathbf{b} \\ \text{subject to:} \\ \quad \mathbf{M}\mathbf{b} \leq \mathbf{v}^{e'} \\ \quad \mathbf{b} \geq \mathbf{0} . \end{array} \right.$$

$$(\text{D-CF}^{e'}) \quad \left\{ \begin{array}{l} \text{Find } \mathbf{y} \in \mathbb{R}^m \\ \text{minimising } \mathbf{y} \cdot \mathbf{v}^{e'} \\ \text{subject to:} \\ \quad \mathbf{M}^T \mathbf{y} \geq \mathbf{1} \\ \quad \mathbf{y} \geq \mathbf{0} . \end{array} \right.$$

We denote the positive cone in  $\mathbb{R}^d$  by  $\mathbb{R}^{d+}$  for an integer  $d$ . The Lagrangian function [67]  $\mathbb{R}^{n+} \times \mathbb{R}^{m+} \rightarrow \mathbb{R}$  for (P-CF<sup>e</sup>) and (D-CF<sup>e</sup>) is, for  $\mathbf{b} \in \mathbb{R}^{n+}$  and  $\mathbf{y} \in \mathbb{R}^{m+}$ :

$$\mathcal{L}(\mathbf{b}, \mathbf{y}) = \mathbf{1} \cdot \mathbf{b} - \mathbf{y}^T(\mathbf{M}\mathbf{b} - \mathbf{v}^e) = \mathbf{y} \cdot \mathbf{v}^e - \mathbf{b}^T(\mathbf{M}^T\mathbf{y} - \mathbf{1}) . \quad (\text{B.46})$$

The primal program (P-CF<sup>e</sup>) indeed corresponds to

$$\sup_{\mathbf{b} \in \mathbb{R}^{n+}} \inf_{\mathbf{y} \in \mathbb{R}^{m+}} \mathcal{L}(\mathbf{b}, \mathbf{y}) , \quad (\text{B.47})$$

while the dual program (D-CF<sup>e</sup>) to

$$\inf_{\mathbf{y} \in \mathbb{R}^{m+}} \sup_{\mathbf{b} \in \mathbb{R}^{n+}} \mathcal{L}(\mathbf{b}, \mathbf{y}) . \quad (\text{B.48})$$

We denote  $f$  the dual Lagrangian function, that is:

$$\begin{aligned} f : \mathbb{R}^{m+} &\longrightarrow \mathbb{R} \\ \mathbf{y} &\longmapsto \sup_{\mathbf{b} \in \mathbb{R}^{n+}} \left[ \mathbf{1} \cdot \mathbf{b} - \mathbf{y}^T(\mathbf{M}\mathbf{b} - \mathbf{v}^e) \right] , \end{aligned} \quad (\text{B.49})$$

which is well-defined under the constraint  $\mathbf{M}^T\mathbf{y} - \mathbf{1} \geq 0$ . The relation  $\geq$  has to be understood as a cone constraint i.e. the fact the all coefficients are positive.

Of course, we can have a similar treatment for  $e'$ . This yields the dual Lagrangian function:

$$\begin{aligned} f' : \mathbb{R}^{m+} &\longrightarrow \mathbb{R} \\ \mathbf{y} &\longmapsto \sup_{\mathbf{b} \in \mathbb{R}^{n+}} \left[ \mathbf{1} \cdot \mathbf{b} - \mathbf{y}^T(\mathbf{M}\mathbf{b} - \mathbf{v}^{e'}) \right] . \end{aligned} \quad (\text{B.50})$$

Importantly, note that (P-CF<sup>e</sup>) and (D-CF<sup>e</sup>) (resp. (P-CF<sup>e'</sup>) and (D-CF<sup>e'</sup>)) are strongly dual (i.e. they have the same optimal value) and they both have optimal solutions (since they are finite-dimensional linear programs with a finite value and they have interior points i.e. points that strictly satisfy all constraints [67]).

Let  $\mathbf{y}^* \in \mathbb{R}^{m+}$  be an optimal solution of (D-CF<sup>e</sup>) and let  $\mathbf{b}' \in \mathbb{R}^{n+}$  be a feasible point of (P-CF<sup>e'</sup>). Then:

$$\text{NCF}(e) = f(\mathbf{y}^*) \quad (\text{B.51})$$

$$\geq \mathbf{1} \cdot \mathbf{b}' - \mathbf{y}^{*T}(\mathbf{M}\mathbf{b}' - \mathbf{v}^e) \quad (\text{B.52})$$

$$\geq \mathbf{1} \cdot \mathbf{b}' + \mathbf{y}^* \cdot \mathbf{v}^\omega , \quad (\text{B.53})$$

where Eq. (B.51) follows from strong duality; Eq. (B.52) follows from the definition of  $f$  with the supremum on  $\mathbf{b} \in \mathbb{R}^{n+}$  in Eq. (B.49); and Eq. (B.53) follows from the fact that  $\mathbf{y}^* \geq 0$  and that if  $\mathbf{b}'$  is feasible for (P-CF<sup>e'</sup>) then  $\mathbf{M}\mathbf{b}' \leq \mathbf{v}^e - \mathbf{v}^\omega$  and thus  $-\mathbf{M}\mathbf{b}' + \mathbf{v}^e \geq \mathbf{v}^\omega$ . Since this holds for every  $\mathbf{b}'$  feasible for (P-CF<sup>e'</sup>) then in particular at an optimal solution  $\mathbf{b}'^*$  for which  $\mathbf{1} \cdot \mathbf{b}'^* = \text{NCF}(e')$ :

$$\text{NCF}(e) \geq \text{NCF}(e') + \mathbf{y}^* \cdot \mathbf{v}^\omega . \quad (\text{B.54})$$

Now, symmetrically, let  $\mathbf{y}'^* \in \mathbb{R}^{m+}$  be an optimal solution of (D-CF<sup>e'</sup>) and let  $\mathbf{b} \in \mathbb{R}^{n+}$  be a feasible point of (P-CF<sup>e</sup>). Then:

$$\text{NCF}(e') = f'(\mathbf{y}'^*) \quad (\text{B.55})$$

$$\geq \mathbf{1} \cdot \mathbf{b} - \mathbf{y}'^{*T}(\mathbf{M}\mathbf{b} - \mathbf{v}^{e'}) \quad (\text{B.56})$$

$$\geq \mathbf{1} \cdot \mathbf{b} - \mathbf{y}'^* \cdot \mathbf{v}^\omega , \quad (\text{B.57})$$

where Eq. (B.57) follows from the facts that  $\mathbf{y}^{/*} \geq 0$  and that if  $\mathbf{b}$  is a feasible point for the program (P-CF<sup>e</sup>) then  $M\mathbf{b} \leq \mathbf{v}^{e'} + \mathbf{v}^\omega$  so that  $-M\mathbf{b} + \mathbf{v}^{e'} \geq -\mathbf{v}^\omega$ . In particular, at an optimal solution  $\mathbf{b}^* \in \mathbb{R}^{n+}$  for (P-CF<sup>e</sup>) for which  $\mathbf{1} \cdot \mathbf{b}^* = \text{NCF}(e)$ :

$$\text{NCF}(e') \geq \text{NCF}(e) - \mathbf{y}^{/*} \cdot \mathbf{v}^\omega . \quad (\text{B.58})$$

From Eq. (B.54) and Eq. (B.58), we have that:

$$\text{NCF}(e) + \mathbf{y}^* \cdot \mathbf{v}^{\omega-} \geq \text{NCF}(e') \geq \text{NCF}(e) - \mathbf{y}^{/*} \cdot \mathbf{v}^{\omega+} , \quad (\text{B.59})$$

where we recall that  $\mathbf{v}^{\omega+}$  is the positive part of  $\mathbf{v}^\omega$  and  $\mathbf{v}^{\omega-}$  is negative part.

Now having a close inspection at the problem (D-CF<sup>e</sup>) (resp. (D-CF<sup>e'</sup>)) gives that  $\mathbf{1} \in \mathbb{R}^{m+}$  is necessarily an upper bound for  $\mathbf{y}^*$  (resp.  $\mathbf{y}^{/*}$ ) in the convex cone of positive vectors, i.e.  $\mathbf{1} \geq \mathbf{y}^*$  (resp.  $\mathbf{1} \geq \mathbf{y}^{/*}$ ). Indeed,  $\mathbf{1}$  trivially satisfies all the constraints and gives a weight of 1 to every component in a context. On the other hand, the optimal solution of (D-CF<sup>e</sup>) (resp. (D-CF<sup>e'</sup>)) will give, at worst, a weight of 1 to one specific component in every context (it needs to satisfy the constraints  $M^T \mathbf{y} \geq \mathbf{1}$  and  $\mathbf{y} \geq 0$  while minimising  $\mathbf{y} \cdot \mathbf{v}^{e'}$ ).

Then, because of Eq. (B.45):

$$\mathbf{y}^* \cdot \mathbf{v}_-^\omega \leq \mathbf{1} \cdot \mathbf{v}_-^\omega \leq |\mathcal{M}|\varepsilon \quad (\text{B.60})$$

$$-\mathbf{y}^{/*} \cdot \mathbf{v}_+^\omega \geq -\mathbf{1} \cdot \mathbf{v}_+^\omega \geq -|\mathcal{M}|\varepsilon , \quad (\text{B.61})$$

so that:

$$\text{NCF}(e) + |\mathcal{M}|\varepsilon \geq \text{NCF}(e') \geq \text{NCF}(e) - |\mathcal{M}|\varepsilon . \quad (\text{B.62})$$

Thus:

$$|\text{NCF}(e) - \text{NCF}(e')| \leq |\mathcal{M}|\varepsilon \quad (\text{B.63})$$

$$|\text{CF}(e) - \text{CF}(e')| \leq |\mathcal{M}|\varepsilon . \quad (\text{B.64})$$

□

## C Proofs of Lemma 1 and Theorem 2

**Lemma 3.** *Let  $\langle \{\lambda\}, \delta_\lambda, (h^\lambda) \rangle$  be a  $(1 - \eta_\lambda)$  and  $(1 - \sigma_\lambda)$  HVM. Then, if  $2\eta_\lambda + \sigma_\lambda < 1$ , the contextual fraction of each hidden variable behaviour satisfies  $\text{CF}(h^\lambda) \leq \eta_\lambda$ .*

*Proof.* To prove Lemma 3 we will decompose  $h^\lambda$  according to its deterministic and non-deterministic part optimally:

$$\exists h_{\text{OD}}^\lambda, h'^\lambda \quad \text{s.t.} \quad h^\lambda = (1 - \eta_\lambda^*)h_{\text{OD}}^\lambda + \eta_\lambda^*h'^\lambda . \quad (\text{C.65})$$

Decomposing optimally implies that  $\eta_\lambda^*$  is irreducible, hence any other decomposition will imply  $\eta_\lambda > \eta_\lambda^*$ .

The proof is divided into two parts. Firstly, when  $h_{\text{OD}}^\lambda$  is parameter independent, the bound is trivially found. Secondly when  $h_{\text{OD}}^\lambda$  is parameter dependent, it is not possible to respect the condition  $\sigma_\lambda + 2\eta_\lambda < 1$ .

Case  $h_{\text{OD}}^\lambda$  is parameter independent.

This case follows from the convexity of the contextual fraction using Eq. (C.65). Since  $h_{\text{OD}}^\lambda$  is parameter independent and deterministic, it must belong to the noncontextual polytope [5, Prop. 3.1 and Th. 8.1], thus its noncontextual fraction is 0. On the other hand the contextual fraction of any model is upper bounded by 1 thus:

$$\begin{aligned} \text{CF}(h^\lambda) &= \text{CF}\left((1 - \eta_\lambda)h_{\text{OD}}^\lambda + \eta_\lambda h'^\lambda\right) \\ &\leq (1 - \eta_\lambda)\text{CF}(h_{\text{OD}}^\lambda) + \eta_\lambda \text{CF}(h'^\lambda) \\ &\leq \eta_\lambda \text{CF}(h'^\lambda) \\ &\leq \eta_\lambda . \end{aligned} \tag{C.66}$$

Again, this is the tightest bound, since we assumed that Eq. (C.65) is optimally decomposed.

Case  $h_{\text{OD}}^\lambda$  is parameter dependent.

In this case we can take  $\eta_\lambda < 0.5$  for otherwise  $2\eta_\lambda + \sigma_\lambda \geq 1$ . Thus if our hidden variable behaviour is of the form (C.65), we will prove by contradiction that it is not possible to satisfy  $2\eta_\lambda + \sigma_\lambda < 1$ . To do so, we rely on the fact that the signalling fraction is lower bounded by the maximum incompatibility of marginals (see Lemma 2) and we prove the following lemma:

**Lemma 4.** *For a decomposition of  $h^\lambda$  of the form  $(1 - \eta_\lambda)h_{\text{OD}}^\lambda + \eta_\lambda h'^\lambda$  (see Eq. (C.65)) where  $h_{\text{OD}}^\lambda$  is parameter dependent and  $\eta_\lambda < 0.5$ , we have  $\text{MIM}(h^\lambda) \geq 1 - 2\eta_\lambda$ .*

*Proof.* Let us decompose further Eq. (C.65):

$$\begin{aligned} h^\lambda &= (1 - 2\eta_\lambda)h_{\text{OD}}^\lambda + 2\eta_\lambda(1/2 h'^\lambda + 1/2 h_{\text{OD}}^\lambda) \\ &= (1 - 2\eta_\lambda)h_{\text{OD}}^\lambda + 2\eta_\lambda h''^\lambda . \end{aligned} \tag{C.67}$$

Since  $h_{\text{OD}}^\lambda$  is deterministic and parameter dependent, we know that  $\text{MIM}(h_{\text{OD}}^\lambda) = 1$ , which implies:

$$\exists C, C' \in \mathcal{M}, t \in O_{CnC'}, \quad h_{\text{OD},C}^\lambda|_{CnC'}(t) - h_{\text{OD},C'}^\lambda|_{CnC'}(t) = 1 . \tag{C.68}$$

Then because  $h_C'^\lambda|_{CnC'}(t) - h_{C'}^\lambda|_{CnC'}(t) \geq -1$ :

$$1/2(h_{\text{OD},C}^\lambda|_{CnC'}(t) - h_{\text{OD},C'}^\lambda|_{CnC'}(t)) + 1/2(h_C'^\lambda|_{CnC'}(t) - h_{C'}^\lambda|_{CnC'}(t)) \geq 0 . \tag{C.69}$$

That is:

$$h_C''^\lambda|_{CnC'}(t) - h_{C'}''^\lambda|_{CnC'}(t) \geq 0 . \tag{C.70}$$

Extending to the general case of Eq. (C.67) leads to:

$$\begin{aligned} h_C^\lambda|_{CnC'}(t) - h_{C'}^\lambda|_{CnC'}(t) &= \\ &= (1 - 2\eta_\lambda)(h_{\text{OD},C}^\lambda|_{CnC'}(t) - h_{\text{OD},C'}^\lambda|_{CnC'}(t)) \\ &\quad + 2\eta_\lambda(h_C''^\lambda|_{CnC'}(t) - h_{C'}''^\lambda|_{CnC'}(t)) \\ &\geq 1 - 2\eta_\lambda , \end{aligned} \tag{C.71}$$

where the last line of Eq. (C.71) is given by equations (C.68) and (C.70). By definition of MIM:

$$\text{MIM}(h^\lambda) \geq 1 - 2\eta_\lambda . \quad (\text{C.72})$$

□

Finally combining Lemma 2 and Lemma 4 we obtain:

$$\begin{aligned} \sigma_\lambda &\geq \text{SF}(h^\lambda) \\ &\geq \text{MIM}(h^\lambda) \\ &\geq 1 - 2\eta_\lambda \\ &\implies \sigma_\lambda + 2\eta_\lambda \geq 1 . \end{aligned}$$

thus we obtain a contradiction with the assumption. □

**Theorem 4.** Let  $e = \sum_\lambda p(\lambda)h^\lambda$  a behaviour realisable by a  $(1 - \sigma)$  PI and  $(1 - \eta)$  OD HVM  $\langle \Lambda, p, (h^\lambda)_{\lambda \in \Lambda} \rangle$  such that  $\sigma + 2\eta < 1$ . Then its contextual fraction is bounded by above:  $\text{CF}(e) \leq \eta$ .

*Proof.* By definition  $\sigma_\lambda + 2\eta_\lambda < \sigma + 2\eta$ , thus by using the convexity of the contextual fraction and applying Lemma 3:

$$\begin{aligned} \text{CF}(e) &\leq \sum_\lambda p(\lambda)\text{CF}(h^\lambda) \\ &\leq \sum_\lambda p(\lambda)\eta_\lambda \\ &\leq \eta \sum_\lambda p(\lambda) \\ &\leq \eta . \end{aligned} \quad (\text{C.73})$$

□

## D High signalling and unsharpness

We rarely expect the case  $\sigma_\lambda + 2\eta_\lambda \geq 1$  to be interesting. From an experimental point of view, considering the departure from the quantum set of correlations to be due to noise, it would mean that the experiment is very noisy. Nonetheless we have a look at this case here and prove for the  $n$ -cycle scenario [46] that we can find a hidden variable model for which  $\sigma_\lambda + 2\eta_\lambda \geq 1$  and  $\text{CF}(h^\lambda) = 1$ . Since the contextual fraction must be upper-bounded by 1, this means that all the contextuality that is witnessed can be attributed to noise.

**Lemma 5.** Let a hidden variable behaviour  $h^\lambda$  be decomposed as:

$$h^\lambda = (1 - \beta)h_{\text{OD}}^\lambda + \beta h'^\lambda , \quad (\text{D.74})$$

where we assume that  $h'^\lambda$  does not contain  $h_{\text{OD}}^\lambda$ . If  $\beta < 0.5$  we have that  $\beta = \eta_\lambda^*$  with  $\eta_\lambda^*$  the irreducible weight on the nondeterministic behaviour.

*Proof.* The closeness of a point to a behaviour is given by the maximal weight one can put on the point in a convex decomposition that explains the behaviour. In other words one can define the closeness  $\rho$  between a deterministic point  $h_{\text{det}}^\lambda \neq h_{\text{OD}}^\lambda$  and the given behaviour  $h^\lambda$  with:

$$\rho = \sup \left\{ r \mid r h_{\text{det}}^\lambda \leq h^\lambda, r \in [0, 1], \text{ for any } h_{\text{det}}^\lambda \neq h_{\text{OD}}^\lambda \right\} . \quad (\text{D.75})$$

Our assumption leads to  $\rho \leq \beta \leq 0.5$ . To show this, by contradiction, if  $\rho > \beta$  then it means that:

$$\exists h_{det}^\lambda \neq h_{OD}^\lambda, \rho h_{det}^\lambda \leq h^\lambda, \quad \rho > \beta.$$

But since  $h_{det}^\lambda \neq h_{OD}^\lambda$  it means that at least for one context they don't have the same support and thus:

$$\exists C \in \mathcal{M}, t \neq t' \in O, h_{OD,C}^\lambda(t) = h_{det,C}^\lambda(t') = 1. \quad (\text{D.76})$$

This implies that  $h_C^\lambda(t) \geq 1 - \beta$  and  $h_C^\lambda(t') \geq \rho$  so summing over this context:

$$\begin{aligned} \sum_{s \in O_C} h_C^\lambda(s) &\geq (1 - \beta) + \rho \\ &> 1, \end{aligned} \quad (\text{D.77})$$

where the last inequality comes from the assumption that  $\rho > \beta$ . This contradicts the normalisation of the hidden variable model, thus we have proven that there does not exist  $h_{det}^\lambda$  that is closer to  $h^\lambda$ . In other words, writing the nondeterministic maximal weight yields

$$h^\lambda = (1 - \eta_\lambda^*) h_{OD}^\lambda + \eta_\lambda^* h'^\lambda \quad (\text{D.78})$$

and since  $h_{OD}^\lambda$  is the closest point it corresponds to the best decomposition.  $\square$

**Lemma 6.** *For the  $n$ -cycle scenario, it is possible to find a hidden variable model such that  $\sigma_\lambda + 2\eta_\lambda = 1$  and  $\text{CF}(h^\lambda) = 1$ .*

*Proof.* Let us define a hidden variable model:

$$\alpha \in [0, 1], \quad h_{ub}^\lambda = \alpha h_{S_1}^\lambda + (1 - \alpha) h_{S_2}^\lambda, \quad (\text{D.79})$$

where  $h_{S_1}^\lambda$  and  $h_{S_2}^\lambda$  are two deterministic parameter dependent hidden variable behaviours and are chosen such that a uniform mixture of both gives a contextual parameter independent behaviour:

$$h_{\text{CPI}}^\lambda = 1/2 h_{S_1}^\lambda + 1/2 h_{S_2}^\lambda. \quad (\text{D.80})$$

In the  $n$ -cycle scenario it is always possible to find such behaviours [46], in the CHSH scenario [68] this corresponds to the PR-box [43]. By using Lemma 5 on equation D.79 we find that  $\eta_\lambda = \min(\alpha, 1 - \alpha)$ . Let us now define another hidden variable behaviour  $h'_{ub}^\lambda$ :

$$\beta \in [0, 1], \quad h'_{ub}^\lambda = \beta h_{S_1}^\lambda + (1 - \beta) h_{\text{CPI}}^\lambda. \quad (\text{D.81})$$

We know that  $\text{MIM}(h_{S_1}^\lambda) = 1$  and  $\text{MIM}(h_{\text{CPI}}^\lambda) = 0$ . Moreover, in this case:

$$\begin{aligned} \text{MIM}(h'_{ub}^\lambda) &= \beta \text{MIM}(h_{S_1}^\lambda) + (1 - \beta) \text{MIM}(h_{\text{CPI}}^\lambda) \\ &= \beta. \end{aligned} \quad (\text{D.82})$$

Using Lemma 2 we have  $\text{SF}(h'_{ub}^\lambda) \geq \beta$ . But using the convexity of the signalling fraction we find:

$$\begin{aligned} \text{SF}(h'_{ub}^\lambda) &\leq \beta \text{SF}(h_{S_1}^\lambda) + (1 - \beta) \text{SF}(h_{\text{CPI}}^\lambda) \\ &\leq \beta. \end{aligned} \quad (\text{D.83})$$

Then  $\text{SF}(h_{ub}^\lambda) = \beta$ . Converting Eq. D.81, by inserting Eq. D.80:

$$\begin{aligned} h_{ub}^\lambda &= \beta h_{S_1}^\lambda + \beta(1/2h_{S_1}^\lambda + 1/2h_{S_2}^\lambda) \\ &= \frac{1}{2}(1 + \beta)h_{S_1}^\lambda + \frac{1}{2}(1 - \beta)h_{S_2}^\lambda \\ &= \alpha h_{S_1}^\lambda + (1 - \alpha)h_{S_2}^\lambda, \quad \forall \alpha \in [1/2, 1], \end{aligned}$$

where  $\alpha$  is related to  $\beta$  by  $\alpha = \frac{1}{2}(1 + \beta)$ . Again  $\eta_\lambda^* = 1 - \alpha$  due to Lemma 5:

$$\begin{aligned} \eta_\lambda^* &= 1 - \alpha \\ \eta_\lambda^* &= 1 - \frac{1}{2}(1 + \beta) \\ \eta_\lambda^* &= 1 - \frac{1}{2}(1 + \sigma_\lambda^*) \\ 2\eta_\lambda^* &= 1 - \sigma_\lambda^* \\ \sigma_\lambda^* &= 1 - 2\eta_\lambda^*. \end{aligned}$$

Hence we have found a hidden variable model  $h^\lambda$  such that  $\sigma_\lambda^* + 2\eta_\lambda^* = 1$ .

We now prove that  $\text{CF}(h_{ub}^\lambda) = 1$ . This proof is based on the fact that  $\text{CF}(h_{CPI}^\lambda) = 1$ , thus by decomposing it we obtain (according to Eq. D.80):

$$\begin{aligned} h_{CPI}^\lambda &= 1/2h_{S_1}^\lambda + 1/2h_{S_2}^\lambda \\ &= \frac{\alpha}{2\alpha}h_{S_1}^\lambda + \frac{1}{2}h_{S_2}^\lambda + \frac{1 - \alpha}{2\alpha}h_{S_2}^\lambda - \frac{1 - \alpha}{2\alpha}h_{S_2}^\lambda \\ &= \frac{1}{2\alpha}h_{ub}^\lambda + \left(\frac{1}{2} - \frac{1 - \alpha}{2\alpha}\right)h_{S_2}^\lambda. \end{aligned} \tag{D.84}$$

Thus by convexity of the contextual fraction one has :

$$\begin{aligned} \text{CF}(h_{CPI}^\lambda) &\leq \frac{1}{2\alpha}\text{CF}(h_{ub}^\lambda) + \left(\frac{1}{2} - \frac{1 - \alpha}{2\alpha}\right)\text{CF}(h_{S_2}^\lambda) \\ 1 &\leq \frac{1}{2\alpha}\text{CF}(h_{ub}^\lambda) + \left(\frac{1}{2} - \frac{1 - \alpha}{2\alpha}\right) \\ &\implies \text{CF}(h_{ub}^\lambda) = 1, \end{aligned} \tag{D.85}$$

where the decomposition is valid only if  $\alpha \geq 0.5$ . For the case  $\alpha \leq 0.5$ , one must use  $h_{S_1}^\lambda$  instead of  $h_{S_2}^\lambda$  in Eq. D.84. Finally, combining all of the above one gets a hidden variable model with  $\sigma_\lambda^* + 2\eta_\lambda^* = 1$  and  $\text{CF}(h_{ub}^\lambda) = 1$ .  $\square$