Mathematical tools for characterising contextual quantum advantage

Vinicius Pretti Rossi

Thesis submitted for the degree of
Doctor of Philosophy at
The University of Gdańsk

Supervisor: dr hab. Ana Belén Sainz, prof UG

International Centre for Theory of Quantum Technologies
University of Gdańsk

Acknowledgments

First and foremost, I would like to thank my supervisor Ana Belén Sainz for her endless support and guidance. I will be always grateful for your presence whenever I needed, and for all the pubs and cafes we got to share. I also thank John Selby who has become a sort of co-supervisor in this journey, and Luca for bringing so much light into his parents' lives. I thank Beata, David, Baldi, Marco, Paulo, Sina, Matthias, Robin, and the rest of FUQT current and past members for all the foundations meetings and pubs, conferences, and paperwork we got to share. A special thanks to my collaborators David, Matty, Baldi, and Amanda, for being the brilliant minds you are. This thesis was only possible thanks to you. To my friends from ICTQT who endured the worst of pandemics and of the doctoral school with me: Chithra, Giuseppe, Sumit, Ekta, Tushita, Gerardo, and the ICTQT staff, my sincere thanks.

I especially thank my parents, Edilaine and Itamar, for their full support even when this meant that they were losing something precious — you taught me that I can handle anything. Immense thanks to Maciek, my journey partner, for becoming my home — you taught me that I don't have to handle everything. To my grandparents, my loving brother and his wife, and my friends from Brazil who were with me when I so badly needed warmth. I dedicate this thesis to all of you.

This work was supported by the Foundation for Polish Science (International Research Agenda Programme project, International Centre for Theory of Quantum Technologies, Grant No. 2018/MAB/5, co-financed by the European Union within the Smart Growth Operational program).

Abstract

Quantum technologies and protocols yielded an unprecedented boost on communication, computation and information processing. However, optimising performance in tasks with quantum resources requires an ever-increasing knowledge of what the nonclassical features of quantum theory are. Research in quantum foundations in the past decade has identified generalised contextuality as one of the best-motivated notions of nonclassicality available, since it simultaneously provides the quantum advantage in many computation, communication and information processing tasks, subsumes or relates to a wide family of other signatures of nonclassicality, and has a reasonable philosophical motivation from Leibniz's assumption of the identity of indiscernibles.

Assessing generalised noncontextuality however is not straightforward, since it usually relies on the explicit verification of the existence of a noncontextual ontological model for the scenario under investigation. This thesis provides a numerical tool for assessing generalised contextuality in prepare-and-measure scenarios, and applies it to different quantum protocols for which contextuality is known to be a resource to explore how more practical quantities and methods based on contextuality can help assess the resourcefulness of these scenarios for communication tasks.

We start by motivating the adoption of generalised contextuality as a notion of nonclassicality by exploring how its device-independent counterpart is not sufficient to draw statements about the realisability of Einstein-Podolsky-Rosen assemblages in Ref. [First Paper]. In particular, we show that if we define a physical principle based on device-independent contextuality, we cannot fully characterise the EPR assemblages that satisfy this principle without making assumptions about the underlying quantum system.

We then shift to the Generalised Probabilistic Theory framework and to the main goal of this thesis in Ref. [Second Paper], where we introduce a linear program for assessing contextuality in prepare-and-measure scenarios. The program estimates the robustness of contextuality to depolarising noise, i.e., how much partial depolarising noise is necessary for a noncontextual ontological model to exist, and provides the ontological model for the depolarised scenario. We also provide an implementation of this program in Mathematica and give some examples of quantum and postquantum scenarios that display or not generalised contextuality.

Finally, we apply this tool to quantum protocols in which contextuality is known to be a resource. In Ref. [Third Paper], we leverage the code to investigate the relation between generalised contextuality and coherence, and show that there are proofs of contextuality that are maximally robust to dephasing noise and that require a vanishing amount of coherence in the states and measurements. We then provide an importation of this linear program to Python in Ref. [Repository]. We employ this new implementation to investigate how well robustness of contextuality to depolarising and to dephasing noise quantify the resourcefulness of scenarios for parity-oblivious multiplexing tasks in Ref. [Preprint]. We conclude that robustness to depolarising noise, as well as robustness to dephasing minimised over key axes, are good quantifiers of the quantum advantage in this task.

Abstrakt

Technologie i protokoły kwantowe zapewniły bezprecedensowy rozwój w dziedzinie komunikacji, obliczeń i przetwarzania informacji. Jednakże, optymalizacja wydajności w zadaniach wykorzystujących zasoby kwantowe wymaga stale rosnącej wiedzy na temat nieklasycznych cech teorii kwantowej. Badania nad podstawami teorii kwantowej przeprowadzone w ostatniej dekadzie wykazały, że uogólniona kontekstualność jest jedną z najlepiej umotywowanych dostępnych koncepcji nieklasyczności, ponieważ jednocześnie zapewnia ona przewagę kwantową w wielu zadaniach obliczeniowych, komunikacyjnych i przetwarzania informacji, obejmuje lub jest związana z wieloma innymi koncepcjami niekasyczności oraz ma rozsądną motywację filozoficzną wynikającą z założenia Leibniza o tożsamości tego, co nierozróżnialne.

Detekcja uogólnionej niekontekstualności nie jest jednak prosta, ponieważ zwykle opiera się na weryfikacji istnienia niekontekstualnego modelu ontologicznego dla badanego scenariusza. Niniejsza rozprawa przedstawia narzędzie numeryczne do detekcji uogólnionej kontekstualności w protokołach *prepare-and-measure* oraz pokazuje jego zastosowanie na różnych protokołach kwantowych, w przypadku których wiadomo, że kontekstualność jest zasobem, w celu zbadania, w jaki sposób bardziej praktyczne wielkości fizyczne i metody oparte na uogólnionej kontekstualności mogą pomóc ocenić przydatność tych protokołów w zadaniach komunikacyjnych.

Rozprawa zaczyna się od umotywowania przyjęcia uogólnionej kontekstualności jako koncepcji nieklasyczności poprzez zbadanie, w jaki sposób wersja *device-independent* kontekstualności nie jest wystarczająca do stwierdzenia kwantowej realizacji asamblaży Einsteina-Podolsky'ego-Rosena w pracy [First Paper]. W szczególności pokazujemy, że jeśli zdefiniujemy fizyczną zasadę makroskopowej niekontekstualności w oparciu o wersje *device-independent* kontekstualności, nie będziemy w stanie w pełni scharakteryzować asamblaży EPR, które spełniają tę zasadę, bez przyjęcia założeń dotyczących układu kwantowego.

Następnie przechodzimy do formalizmu uogólnionej teorii probabilistycznej i do głównego celu tej rozprawy w pracy [Second Paper], gdzie wprowadzamy liniowy program do detekcji kontekstualności w protokołach *prepare-and-measure*. Program szacuje wytrzymałość kontekstualności na szum depolaryzacyjny, tj. ile częściowego szumu depolaryzującego jest konieczne, aby istniał niekontekstualny model ontologiczny, i pokazuje model ontologiczny dla zasobu depolaryzowanego. Przedstawiamy również implementację tego programu w programie Mathematica i podajemy kilka przykładów zasobów kwantowych i postkwantowych, które wykazują lub nie uogólnioną kontekstualność.

W ostatnim rozdziale rozprawy, stosujemy ten program do protokołów kwantowych, w których wiadomo, że kontekstualność jest zasobem. W pracy [Third Paper] wykorzystujemy program liniowy do badania związku między uogólnioną kontekstualnością a koherencją i pokazujemy, że istnieją dowody kontekstualności, które są maksymalnie odporne na szum fazowy i które wymagają zanikającej ilości koherencji w stanach i pomiarach. Następnie zapewniamy tłumaczymy ten program liniowy na język Python w pracy [Repository]. Wykorzystujemy tę nową implementację, aby zbadać, jak dobrze odporność kontekstualności na depolaryzację i szum fazowy określa ilościowo wartość zasobów w zadaniach *parity-oblivious multiplexing* w pracy [Preprint]. Pokazujemy, że odporność na szum depolaryzujący oraz szum fazowy zminimalizowany w głównych osiach, dobrze ilościowo określają przewagę kwantową w tym zadaniu.

Publications included in the dissertation

First Paper *V. P. Rossi*, M. J. Hoban, A. B. Sainz, *On characterising assemblages in Einstein-Podolsky-Rosen assemblages. J. Phys. A: Mat. Theo.* **55** 264002 (2022).

Second Paper J. H. Selby, E. Wolfe, D. Schmid, A. B. Sainz, *V. P. Rossi*, Linear Program for Testing Nonclassicality and an Open-Source Implementation. *Phys. Rev. Lett.* **132**, 050202 (2024).

Third Paper *V. P. Rossi*, D. Schmid, J. H. Selby, A. B. Sainz, Contextuality with vanishing coherence and maximal robustness to dephasing. *Phys. Rev. A* **108**, 032213 (2023).

Repository P. J. Cavalcanti, *V. P. Rossi*, J. H. Selby, A. B. Sainz, `SimplexEmbeddingGPT`: Simplex Embedding for Generalised Probabilistic Theory (GPT) Fragments. GitHub Repository, https://github.com/pjcavalcanti/SimplexEmbeddingGPT (2023).

Preprint A. M. Fonseca, *V. P. Rossi*, R. D. Baldijão, J. H. Selby, A. B. Sainz, Robustness of contextuality under different types of noise as quantifiers for parity-oblivious multiplexing tasks. *arXiv preprint*, arXiv:2406.12773v1 [quant-ph] (2024).

Other publications

OP1 *V. P. Rossi*, D. O. Soares-Pinto, Wigner's friend and the quasi-ideal clock. *Phys. Rev. A* **103** 052206 (2021).

OP2 *V. P. Rossi*, R. P. Bonini, A. M. Gonçalves, A. J. Gualdi, J. A. Eiras, F. L. Zabotto, Silicon substrate orientation influence on structural and magnetic properties of $BaFe_{12}O_{19}$ thin films obtained by RF magneton sputtering. *J. Magn. Magn. Mater.* **504** 166705 (2020).

Contents

Acknowledgments **iii**

Abstract **v**

Abstrakt **vii**

Publications included in the dissertation **ix**

Other publications **xi**

1 Introduction **1**

2 Preliminaries **5**
2.1 Generalised contextuality 5
2.1.1 Operational theories and quotienting 5
2.1.2 Ontological models and noncontextuality 6
2.2 The hypergraph approach to contextuality 8
2.3 Contextuality in the GPT framework 9
2.3.1 GPTs, GPT fragments and accessible GPT fragments 9
2.3.2 Simplex embedding 11
2.4 Summary 12

3 Summary of dissertation **15**
3.1 *On characterising assemblages in Einstein-Podolsky-Rosen scenarios* [First Paper] 15
3.1.1 EPR scenarios and realisability 15
3.1.2 EPR assemblages as correlations 16
3.1.3 Macroscopic assemblages and macroscopic non-contextuality 17
3.2 *Linear program for testing nonclassicality and an open-source implementation* [Second Paper] 18
3.2.1 Positive-cone facets 19
3.2.2 Linear program for simplicial cone embedding 20
3.2.3 Bound on the dimension of the simplicial space 21
3.2.4 Examples 22
3.3 *Contextuality with vanishing coherence and maximal robustness to dephasing* [Third paper] 23
3.3.1 Motivation – contextuality implies coherence 23
3.3.2 Family of coherent prepare-and-measure scenarios 24
3.3.3 Contextuality with vanishing coherence 25
3.3.4 Proof of contextuality maximally robust to dephasing noise 27

3.4 *A Python importation of the linear program from Ref. [Second Paper]* [Repository] . . . 28
3.5 *Robustness of contextuality under different types of noise as quantifiers for parity-oblivious multiplexing tasks* [preprint] . . . 29
3.5.1 Modification of the program . . . 29
3.5.2 Parity-oblivious multiplexing . . . 30
3.5.3 Robustness to depolarising noise in the n-to-1 POM task . . . 31
3.5.4 Robustness to dephasing in the 3-to-1 POM task . . . 32

4 Outlook **33**

List of Figures

3.1 (Left) Robustness of contextuality to depolarising noise vs. success rate for the 3-to-1 POM task. Red curve reproduces Eq. 3.80. (Right) Minimal robustness of contextuality to dephasing noise across the measurements available in the fragment vs. success rate for the 3-to-1 POM task. 32

List of Tables

3.1 Data table for prepare-and-measure scenarios satisfying the equivalence relations introduced previously. 26
3.2 Data table for the partially depolarised prepare-and-measure scenario. . . . 26

Chapter 1

Introduction

In the quest for the development of ever-improving technologies, many are examples implying an advantage in the use of quantum systems over classical ones. From the seminal work of Einstein, Podolsky, and Rosen [1, 2] to Bell's [3, 4] and Kochen-Specker's notions of nonclassicality [5], physicists have developed what are now robust theories for nonlocality [6], contextuality [7], entanglement [8], coherence [9], discord [10], and many other notions of nonclassicality that can provide some form of advantage in one or more tasks. In this multitude of signatures of nonclassicality, Spekkens provided a sound new candidate [11], drawing a clear distinction between the operational prescription on how to compute probabilities and the actual, realist description of the physical properties of the system and its evolution: for Spekkens, a theory is nonclassical when its realist explanation, built upon Bayesian probability theory and Boolean logic, must describe indistinguishable states of affairs as different theoretical entities.

This definition of *generalised contextuality*, firmly grounded on Leibniz's principle of indiscernibles and employed by Einstein in the development of Relativity Theory [12], has been favoured as the best notion of nonclassicality available. Many of the instances of other signatures of nonclassicality trace back as special cases of generalised contextuality or constituting an equivalent counterpart in the right conditions [13–20]. It would however be undermining to adopt a broader notion of nonclassicality if, when turning to practical examples, one would still need to rely on signatures mentioned previously. That is not the case for contextuality: there are many tasks across communication [21–23], computation [24], information processing [25–28], and others [29–35] in which quantum realisations perform better than classical ones have been shown to have contextuality as the source of this advantage, making it philosophically, theoretically and practically interesting.

Some objections towards contextuality have been raised, as discussed in detail in Ref. [36]. A popular one is that contextuality, as opposed to Bell nonlocality, cannot be investigated in a device-independent way. The idea of device-independence is present in quantum phenomena that are allegedly witnessed exclusively by looking at the statistics of the experiment without the need to make any assumption of the underlying physical system. Generalised contextuality is indeed not of this type since establishing what states and measurements are indistinguishable requires one to perform tomographic procedures, which requires knowledge of the properties of the system. But beyond the fact that this so-called device-independence is debatable, a more objective counterargument is that there are many examples of quantum protocols in which device-independent principles do not capture the advantage over classical realisations.

Another aspect justifying the popularity of device-independent approaches is that there are often more resources to work on theoretically. Indeed, the study of non-locality and

Kochen-Specker contextuality is at least half a century old, and there are many numerical tools such as hierarchies of semidefinite programs [37–39] developed for it. Generalised contextuality, on its turn, has accumulated some amount of numerical tools to handle it [40–42], but there is much to do towards implementations that are easy to use and useful for analysing practical scenarios. Although this is much more a matter of the development of a field, it does diverge active research since physicists working on more applied fields might prefer to stick to these other notions of nonclassicality to save resources instead of look for explicit noncontextual ontological models for their experiments.

This thesis targets this objection by introducing an open-source linear program that tests whether prepare-and-measure scenarios can display nonclassicality in the sense of Spekkens and showcasing how theoretical physicists can use it to investigate scenarios with practical applications in communication tasks. We achieve this goal through the three papers, the repository, and the preprint that compose this thesis.

Our first publication addresses the device-dependent objection and serves as yet another example that device-independent notions will only go so far in characterising nonclassicality. In this paper, we investigate Einstein-Podolsky-Rosen (EPR) scenarios [43], an instance of common-cause scenarios in which one of the parties knows everything about the system they hold. We then explore how a version of nonclassicality of device-independent flavour, the hypergraph contextuality, can help us characterise sets of EPR assemblages by asking what types of assemblages conform to the assumption that macroscopically, assemblages must be hypergraph noncontextual. For such, we rely on the fact that EPR assemblages are lists of correlations like any other device-independent scenario, with the caveat that one of the agents knows that they perform a tomographic set of measurements in their lab. As expected, this device-independent principle is not enough to fully characterise the assemblages, and a stronger version of this principle, addressing this additional knowledge, must be imposed to get any meaningful bound.

We then shift our attention to generalised contextuality in our second paper, leveraging tools from generalised probabilistic theories (GPTs) to prove that finding a noncontextual ontological model for a prepare-and-measure scenario is an instance of a linear program. We also provide an open-source implementation of this program in Mathematica that estimates how much depolarising noise must be added to the scenario so that a noncontextual ontological model exists, an operational measure of contextuality that we coin *robustness to depolarisation*. We show simple examples of how to use this code to characterise prepare-and-measure scenarios in stabiliser, quantum, and boxworld theories, also deriving a bound on the maximum dimension the ontic space of a noncontextual ontological model can assume for a given scenario.

In the third paper, we showcase how the program can be used to investigate the relation between contextuality and other notions of nonclassicality. In particular, we explore a family of prepare-and-measure scenarios related to the minimum-error state-discrimination task [32] to investigate the interplay between the existence of contextuality and coherence. Coherence is known to be necessary for proofs of contextuality, and yet there are examples of noncontextual toy theories that display contextuality [44], which might lead to the intuition that there is a sufficient amount of coherence that allows for contextuality. We show that this is not the case, finding examples of prepare-and-measure scenarios for which any amount of coherence in the preparations and measurements is enough for a proof of contextuality. Moreover, we also provide examples of such proofs that are maximally robust to dephasing noise.

In order to further extend the reach of this linear program, we provide an importation of the Mathematica implementation to Python in Ref. [Repository]. This open-source implementation does essentially the same as the original, but more efficiently as is expected from Python, also making this tool more accessible to the community. We also use this new implementation in Ref. [Preprint] to investigate how well robustness of contextuality to depolarisation and to dephasing noise perform in scenarios related to the parity-oblivious multiplexing task. This is done by exploring the relation between the robustness computed by the linear program and the success rate that usually quantifies contextuality in this scenario. We derive an analytical relation between robustness to depolarisation and the success rate for any quantum system and show how the linear program can be used to derive bounds on the maximal robustness to depolarisation that can be achieved for a system. This bound and the analytical relation can be used to estimate the maximal amount of classical bits that can be optimally encoded in the quantum system for this task. Finally, we explore the particular case in which 3 bits are encoded in a qubit, and show that robustness to dephasing with respect to a particular basis is not a good quantifier, but minimising this quantity with respect to key bases that capture the symmetry of the task restores its usefulness.

To convey these findings, I structure this dissertation as follows: Chapter 2 introduces the theoretical background on generalised contextuality, as well as the two frameworks used in our publications to investigate it. Namely, it introduces the hypergraph approach to contextuality, as well as the GPT framework and the notion of simplex embedding that is key to this thesis. Chapter 3 summarises the publications and additional content that compose this thesis. My efforts in this Chapter were focused on providing a support text to help readers while they go through the publications. Finally, in Chapter 4 I comment on some current and near-future research that is an immediate follow-up to this material.

Chapter 2

Preliminaries

Science is the knowledge about the physical world that is built upon objectively testable facts. A naive understanding of this definition might lead to believing that all things scientific can be interacted with in some experiment, but a quick check throughout Physics brings up many concepts that cannot be observed or detected. The common understanding among physicists seems to be that such abstractions can stick around as long as they help *understanding* experimental results, and yet it is often the case that papers will address abstract objects as real-world entities, or that they will confer to objectively testable things some theoretical flexibility that departs from experimental capabilities, especially when it comes to Quantum Theory. In the following sections, I will comment on the distinction between these two levels of reasoning, and how we can define nonclassicality from them. I then introduce the essential tools used in this thesis for investigating contextuality.

2.1 Generalised contextuality

2.1.1 Operational theories and quotienting

An *operational* approach [45–47] to an experiment will concentrate on descriptions that pertain to the laboratory: instructions on how to prepare a sample; how to align mirrors in the optical table or which field generators to turn on; where the single-photon detectors should be placed, and so on. We can reason about experiments – in particular, those that consist of preparing samples and measuring some of their physical properties – in an operational way: consider (i) the set $\mathcal{P}$ of preparation procedures, i.e., instructions on how to prepare the system; (ii) the set $\mathcal{M}$ of measurement procedures, i.e., instructions on how to observe the physical properties of the prepared system; (iii) and K the set of measurement outcomes, i.e., the labels of the detectors for each measurement procedure $M \in \mathcal{M}$. The culminating point of any experiment is the data it outputs, i.e., the frequencies with which a determinate outcome $k \in K$ for a measurement procedure $M \in \mathcal{M}$ happens, given that the system was prepared following the instructions in $P \in \mathcal{P}$. These frequencies are captured in this approach by defining a bilinear map $p : (K \times \mathcal{M}) \times \mathcal{P} \to [0, 1]$, such that $p(k|M, P)$ is interpreted as the probability of outcome k being observed when measurement M was implemented over a system that has undergone a preparation P and we refer to the map p as a *probability rule*. An *operational theory* is an extrapolation of this framework in which the sets $\mathcal{P}$, $\mathcal{M}$ and K contain *all logically possible experimental procedures for a particular system or collection of systems*, despite what is feasible in a laboratory (or in all laboratories of the world with the current technology).

A crucial aspect of this operational approach is that many different laboratory procedures yield the same experimental data. This fact imposes a notion of *operational equivalence* that can be captured by an equivalence relation $\sim$ between procedures: two preparation procedures P and $P' \in \mathcal{P}$ are operationally equivalent when they yield the same statistical data for all possible outcomes of all measurements. Similarly, two measurement outcomes $k|M$ and $k'|M' \in K \times \mathcal{M}$ are operationally equivalent when they yield the same statistics for all possible preparation procedures. Mathematically, this is defined as

$$P \sim P \iff p(k|M,P) = p(k|M,P'), \quad \forall k|M \in K \times \mathcal{M}; \tag{2.1}$$

$$k|M \sim k'|M' \iff p(k|M,P) = p(k'|M',P), \quad \forall P \in \mathcal{P}. \tag{2.2}$$

The equivalence classes, therefore, must be taken with respect to all logically possible procedures, not only those that can be implemented in a particular laboratory. We represent an operational theory by the tuple $(\mathcal{P}, K \times \mathcal{M}, p, \sim)$. Quantum theory is an example of such a theory, in which preparation procedures are represented by density operators; measurement outcomes are represented by POVM elements; the probability rule is given by the Born rule; and the equivalences are naturally captured by the convex-linearity of the Hilbert space.

Since these equivalences play such a strong role in an operational theory, essentially dividing what information can be distinguished via experimental data and what cannot, we can shift our attention to a more compact operational description in which all "excess information" (sometimes called a *context*) is ignored. This *quotiented theory* is given by the tuple $(\mathcal{P}/\sim, (K \times \mathcal{M})/\sim, p)$, where $\mathcal{P}/\sim$ is the set of all equivalence classes $[P]$ such that

$$[P] := \{P' \in \mathcal{P} : P' \sim P\}, \tag{2.3}$$

and $(K \times \mathcal{M})/\sim$ is the set of all equivalence classes $[k|M]$ such that

$$[k|M] := \{k'|M' \in K \times \mathcal{M} : k'|M' \sim k|M\}, \tag{2.4}$$

with the probability rule being preserved, since

$$p(k|M,P) = p([k|M],[P]), \quad \forall k \in K, M \in \mathcal{M}, P \in \mathcal{P}. \tag{2.5}$$

2.1.2 Ontological models and noncontextuality

Despite being useful for making predictions, this operational framework cannot provide any understanding about why a particular preparation influences the likelihood of measurement outcomes, but rather only captures the correlations between them. For a deeper explanation, we would need an *ontological model*, i.e., a theory that prescribes the true physical properties a system has, and which values these properties assume at all times. That is because physicists have built a (fairly reasonable) expectation that physical properties exist and have definite values, independently of being observed or interacted with in an experiment. In such a theory, the laboratory procedures described by the operational theory are a matter of coarse-graining or revealing the values of the system's physical properties.

Mathematically, this means that physical systems must be associated with some form of measurable space Λ, denominated the *ontic space*. Each element $\lambda \in \Lambda$ assigns values for each of the system's physical properties. Preparing a system therefore becomes a probability distribution $\lambda : \Lambda \times \mathcal{P} \to [0,1]$, assigning a probability $\mu(\lambda|P)$ for each $\lambda \in \Lambda$ given a preparation procedure $P \in \mathcal{P}$. The function $\mu(\lambda|P)$ often receives the name of *epistemic state*, since it represents the randomness introduced to the knowledge about the true state of the system by following the procedure P. Detecting outcomes, in its turn, is understood as reading the configuration λ and assigning a probability of the detector $k \in K$ clicking during a measurement $M \in \mathcal{M}$. This is captured through the *response function* $\xi : (K \times \mathcal{M}) \times \Lambda \to [0, 1]$. The probability of observing the outcome $k \in K$ when measurement $M \in \mathcal{M}$ is performed after preparation $P \in \mathcal{P}$ has been implemented is given by coarse-graining over all possible configurations λ of the values of the physical properties, i.e.,

$$p(k|M, P) = \sum_{\lambda \in \Lambda} \xi(k|M, \lambda)\mu(\lambda|P), \quad \forall k \in K, M \in \mathcal{M}, P \in \mathcal{P}. \tag{2.6}$$

Naturally, one would expect that the equivalences captured at the operational level also hold in the ontological model. This is a common procedure in Physics, done several times by Einstein in his work founding relativity theory, and has its roots in Leibniz's principle of the *identity of indiscernibles*, which can be summarised as the expectation that *if there are two distinct states of affairs underlying the same statistical data, then there should be some observation capable of distinguishing them* [12]. This assumption imposes constraints on the ontological model, since

$$P \sim P' \implies \mu(\lambda|P) = \mu(\lambda|P'), \quad \forall \lambda \in \Lambda; \tag{2.7}$$

$$k|M \sim k'|M' \implies \xi(k|M, \lambda) = \xi(k'|M', \lambda), \quad \forall \lambda \in \Lambda. \tag{2.8}$$

In other words, we can also say that it is natural to expect that an ontological model for an operational theory can also explain the quotiented version of that theory, ignoring the contexts altogether. This notion, first formalised by Spekkens [11], is what we call *generalised noncontextuality.* The advantage of this assumption is that it can be easily extended for other scenarios, such as the ones that include transformations or scenarios presenting other causal structures. Most importantly, Spekkens also demonstrates in his seminal work that quantum theory cannot conform to this assumption, even for experiments in a qubit. Therefore, an ontological explanation for the correlations predicted by quantum theory must either depart from the standard formulation of ontological models, which relies on Bayesian probability theory and Boolean logic or give up the natural Leibnizean assumption of the identity of indiscernibles.

Being such a naturally motivated assumption, the impossibility of complying with noncontextuality (which in this work will be referred to as *contextuality*) configures an interesting theoretical device for assessing the nonclassicality of operational theories. But beyond theoretical usefulness, contextuality has practical applications since its manifestation in quantum scenarios can usually be linked to some form of advantage on particular tasks in communication [21–23], computation [24], information processing [25–28], and many other tasks [29–35]. Furthermore, contextuality is known to be related to or subsume many other notions of nonclassicality, in particular, Bell nonlocality [13, 14, 32], Kochen-Specker contextuality [5], the notion of nonclassicality in Quantum Darwnism [15], the detection of anomalous weak values [16], and the notion of macrorealism [17]. Noncontextuality was also proved equivalent to the existence of a non-negative quaisprobability representation [18–20] and, most importantly for this thesis, to the existence of a simplex embedding for the generalised probabilistic theory (GPT) associated to the operational theory [48].

This thesis explores this practical aspect of contextuality by investigating how it relates to other notions of nonclassicality that are resources for quantum advantage in communication tasks. For such, I will leverage two frameworks for investigating contextuality: the *hypergraph-theoretical* approach, employed in Ref. [First Paper] to explore bounds in the quantum realisability of Einstein-Podolsky-Rosen scenarios, and the framework of *generalised probabilistic theories*, which will yield a linear program for testing the contextuality of prepare-and-measure scenarios in Ref. [Second Paper], and to a direct application for minimum-error state-discrimination scenarios in Ref. [Third Paper] and parity-oblivious multiplexing in Ref. [Preprint].

2.2 The hypergraph approach to contextuality

The hypergraph approach [49] stems from a long-termed research program that focuses on the device-independent characterisation of operational correlations, inaugurated by Popescu and Rohrlich [50] and culminating on the successful Navascués-Pironio-Acín hierarchy and the characterisation of many physical principles in the search for a reconstruction of quantum theory [51–56]. Due to this device-independent motivation it cannot describe preparation procedures, and although it identifies equivalent measurement outcomes, it does not prescribe how to evaluate operational equivalences to begin with. That is because by being device-independent, this approach cannot make assumptions about how the underlying physical system should be described, so it cannot produce statements such as "two preparations yield equal probabilities *for all possible measurement outcomes*", since such statements assume what all the possible measurement outcomes are. Therefore, the hypergraph approach assesses a particular case of generalised contextuality but it will miss nonclassicality that relies on some nontrivial equivalence classes.

In this approach, one takes a collection of measurements and outcomes $\{k|M\}_{k\in K,M\in\mathcal{M}}$ and maps it to an *event hypergraph* $H=(V,E)$, in which V is a set of vertices $v\in V$ representing the tuples $(k|M)$ in the scenario, and E is a collection of subsets of V representing all the possible outcomes of a particular preparation and measurement choice M, denominated *hyperedges* $e\in E$. Moreover, it prescribes that two equivalent measurement outcomes $k|M\sim k'|M'$ should be mapped to the same vertex $v\in V$, although it does not define the equivalence relation $\sim$ or how to assess it. Finally, the probability rule p is imbued in the hypergraph as a weighting on each vertex, i.e., each value $p(k|M)$ is mapped to a probability $p(v)$ to the vertex associated to its respective event. Completeness of measurements imply that hyperedges are normalised, i.e.,

$$\sum_{v\in e} p(v)=1,\quad \forall e\in E. \tag{2.9}$$

In this framework, the operational theory is deemed *classical* if there is a set Λ, a normalised probability distribution $q:\Lambda\to[0,1]$, and a truth assignment $p:V\times\Lambda\to\{0,1\}$ such that

$$p(v)=\sum_{\lambda\in\Lambda} q(\lambda)p(v|\lambda),\quad \forall v\in V. \tag{2.10}$$

One can also employ this framework to investigate whether the operational data admits of a quantum realisation, or if it belongs to a beyond-quantum one. This boils down to checking for the existence of a (finite-dimensional) Hilbert space $\mathcal{H}$, a pure state $|\psi\rangle\in\mathcal{H}$ and projective measurement elements $\{\Pi_v\in\mathcal{D}^*(\mathcal{H})\}_{v\in V}$, such that

$$\sum_{v\in e}\Pi_v=\mathbb{1},\quad \forall e\in E; \tag{2.11}$$

$$p(v)=\langle\psi|\Pi_v|\psi\rangle\,,\quad \forall v\in V. \tag{2.12}$$

Finally and most importantly for this thesis, given an hypergraph H, the approach allows for the definition of an hierarchy of certificates that converges to the set of quantum realisable assignments p. Each level of this hierarchy is verifiable through a semidefinite program, and in the case where the operational scenario describes a Bell scenario, Ref. [49] shows that the first level of this hierarchy coincides with the well-known set of *almost quantum correlations* [39]. In particular, the operational data p is said to be $\mathcal{Q}_1$, i.e., it belongs to the first level of this hierarchy, if there is a positive semidefinite matrix Γ of dimension $1 + |V|$, such that for all $u, v \in \emptyset \cup V$, the following are satisfied:

- $\sum_{u \in e} \Gamma_{u,v} = \Gamma_{\emptyset,v}, \quad \forall v \in V$;
- $\sum_{v \in e} \Gamma_{\emptyset,v} = 1$;
- If $u, v \in e$ and $u \neq v$, then $\Gamma_{u,v} = 0$;
- $\Gamma_{v,v} = p(v)$.

In fact, it has been proven that when the hypergraph H is associated to a Bell scenario, then the set of $\mathcal{Q}_1$ correlations becomes equivalent to the almost quantum set of correlations. Furthermore, this level of the hierarchy also bounds the assignments p for an hypergraph H that comply to the assumption of *macroscopic noncontextuality*, a notion of classicality that demands all macroscopic implementations of operational theories to satisfy the assumption of noncontextuality, given a definition of macroscopic implementation. This will be employed in an attempt to bound the set of quantum assemblages in Einstein-Podolsky-Rosen scenarios in Ref. [First Paper]. As we demonstrate, this formulation is not sufficient to draw statements about the underlying assemblages precisely due to the device-independent take it is built upon.

2.3 Contextuality in the GPT framework

2.3.1 GPTs, GPT fragments and accessible GPT fragments

Another framework to describe operational theories is to describe the *generalised probabilistic theory* (GPT) associated with it. GPTs also stem from the reconstruction program, and many successful attempts to reframe the axioms of quantum theory into information-theoretical principles have been made since the first milestone [45, 57, 58].

The GPT associated to an operational theory is again a mapping from the quotiented theory $(\mathcal{P}/\sim, (K \times \mathcal{M})/\sim, p)$ to a finite-dimensional, real vector space V equipped with an inner product $\langle \cdot, \cdot \rangle$. Each equivalence class of preparations $[P]$ is mapped to a vector $s_P \in V$, with Ω being the set of all such vectors. Similarly, each equivalence class of measurement outcomes $[k|M]$ is mapped to a vector $e_{k|M} \in V$, with $\mathcal{E}$ being the set of all such vectors. Furthermore, these sets $(\Omega, \mathcal{E})$ need to satisfy the following properties:

- Ω is a compact, convex set such that $\mathsf{LinSpan}(\Omega) = V$, and such that $0 \notin \mathsf{AffSpan}(\Omega)$, where 0 is the null vector;
- $\mathcal{E}$ is isomorphic to a subset of Ω^*, and both the null vector 0 and the unit vector u are in its affine span;
- The probability rule is captured by the inner product, i.e., $p(k|M, P) = \langle s_P, e_{k|M} \rangle$;
- Ω contains the normalised counterparts of the state vectors, i.e., $s_P \in \Omega \implies \frac{1}{\langle s_P, u \rangle} s_P \in \Omega$;

- Ω and $\mathcal{E}$ are *tomographic* for each other, i.e.,

$$s_P = s_{P'} \iff \langle s_P, e_{k|M} \rangle = \langle s_{P'}, e_{k|M} \rangle\,, \quad \forall e_{k|M} \in \mathcal{E}; \tag{2.13}$$

$$e_{k|M} = e_{k'|M'} \iff \langle s_P, e_{k|M} \rangle = \langle s_P, e_{k'|M'} \rangle\,, \quad \forall s_P \in \Omega. \tag{2.14}$$

The GPT is given by the tuple $(V, \langle \cdot, \cdot \rangle, \Omega, \mathcal{E})$.

It is important to mention that representing states and effects both as vectors in the vector space V and the probability rule as the linear product is not the only way of formulating a GPT. In fact, another very widespread convention is to represent states as vectors in V, while effects are represented by covectors living in the dual space V. The probability in this case is given by the action of an effect over a state, $e(s)$. For all the cases we are interested in this thesis, these two conventions are equivalent. This main text, as well as the main texts of publications Ref. [Second Paper] and Ref. [Third Paper] adopt the convention with the inner product. However, the dual-space convention allows for a very convenient diagrammatic representation that stems from process theories [59], and is employed in the Supplemental Material of Ref. [Second Paper] to derive most of our results.

As an example, one can think of the operational theory of all the possible preparations and measurements to be implemented over a qubit. In this case, equivalence classes of preparation procedures are represented by (possibly subnormalised) density operators $\rho_P \in \mathcal{D}(\mathcal{H})$, while equivalence classes of measurement outcomes are given by POVM elements $M_k \in \mathcal{D}^*(\mathcal{H})$. As it is well-known, the sets $\mathcal{D}(\mathcal{H})$ and $\mathcal{D}^*(\mathcal{H})$, respectively the set of positive operators with trace ≤ 1 and its dual, are isomorphic. By picking an operator basis representation, for instance the Pauli basis, we can write each of these operators as a vector in $\mathbb{R}^4$:

$$\rho_P = \frac{1}{2}(\mathrm{tr}(\rho_P) + \mathrm{tr}(X\rho_P) + \mathrm{tr}(Y\rho_P) + \mathrm{tr}(Z\rho_P)) \mapsto s_P = \frac{1}{2}\begin{pmatrix} \mathrm{tr}(\rho_P) \\ \mathrm{tr}(X\rho_P) \\ \mathrm{tr}(Y\rho_P) \\ \mathrm{tr}(Z\rho_P) \end{pmatrix}, \tag{2.15}$$

and similarly for the POVM elements M_k mapped to effect vectors $e_{k|M}$. The probability rule is simply given by the scalar product, $p(k|M, P) = e_{k|M}^T \cdot s_P$.

In principle, however, operational scenarios will not contain all the possible preparation and measurement procedures in the theory, but rather a small collection of them. In such cases, we can construct the *GPT fragment* associated with this scenario, i.e., the tuple $(\Omega^F, \mathcal{E}^F, V, \langle \cdot, \cdot \rangle)$ such that $\Omega^F \subset \Omega$ and $\mathcal{E}^F \subset \mathcal{E}$. Notice that for GPT fragments, it might be that $\mathsf{LinSpan}(\Omega^F) \neq \mathsf{LinSpan}(\mathcal{E}^F)$, so tomography does not necessarily hold.

Although simple, this description of GPT fragments is not the most natural way to describe experiments. An alternative way is to represent the sets Ω^F and $\mathcal{E}^F$ with respect to the spaces they span, which we refer to as the *accessible GPT fragment* [60]. We define this representation by the projectors $P_\Omega : V \to \mathsf{LinSpan}(\Omega^F)$ and $P_\mathcal{E} : V \to \mathsf{LinSpan}(\mathcal{E}^F)$, such that $\Omega^A = P_\Omega(\Omega^F)$ and $\mathcal{E}^A = P_\mathcal{E}(\mathcal{E}^F)$. We can also return to the original GPT fragment through the inclusion maps $I_\Omega : \mathsf{LinSpan}(\Omega^F) \to V$ and $I_\mathcal{E} : \mathsf{LinSpan}(\mathcal{E}^F) \to V$, which are equivalent to the pseudoinverses of the projectors. What is lost in the accessible GPT fragment is the probability rule, since a notion of inner product between mismatching spanned spaces is lacking. We can avoid this by employing the inclusion maps to return to the vector space V where the inner product is well defined, so that

$$p(k|M, P) = \langle I_\Omega(s_P), I_\mathcal{E}(e_{k|M}) \rangle\,, \quad \forall s_P \in \Omega^A, e_{k|M} \in \mathcal{E}^A. \tag{2.16}$$

The accessible GPT fragment is given by the tuple $(\Omega^A, \mathcal{E}^A, I_\Omega, I_\mathcal{E})$[1]. It is worth mentioning that a GPT is an accessible GPT fragment with trivial inclusion and projection maps.

2.3.2 Simplex embedding

Most importantly for this thesis is the notion of simplex embedding, which defines a notion of nonclassicality for generalised probabilistic theories that includes but is not restricted to quantum realisable ones. An accessible GPT fragment is said to be *simplex-embeddable* [48], or to *admit of a simplex embedding*, if there are:

- a simplex $\Delta \subset \mathbb{R}^n$, i.e., the convex hull of a finite collection of affinely independent vectors $\{\lambda_i\}_{i=1}^n$, with $n \in \mathbb{N}$;
- linear maps $\iota : \mathsf{LinSpan}(\Omega^A) \to \mathbb{R}^n$ and $\kappa : \mathsf{LinSpan}(\mathcal{E}^A) \to \mathbb{R}^n$,

such that

$$\iota(\Omega^A) \subseteq \Delta; \quad \kappa(\mathcal{E}^A) \subseteq \Delta^*, \tag{2.17}$$

and such that the inner product is preserved by the maps ι and κ, i.e.,

$$\langle I_\Omega(s_P), I_\mathcal{E}(e_{k|M})\rangle = \kappa(e_{k|M}) \cdot \iota(s_P). \tag{2.18}$$

Evidently, the notion can also be defined for GPT fragments and full GPTs in a straightforward manner. Furthermore, the existence of a simplex embedding for a GPT implies the existence of a simplex embedding for any fragment or accessible fragment subsumed by this GPT. Conversely, assessing the impossibility of a simplex embedding for a fragment or accessible fragment implies that the full GPT will also not be simplex-embeddable.

Notice that differently from the hypergraph approach, the GPT framework *assumes* a real vector space in which those states and effects live in, and so statements of the form "two preparations yield the same probabilities for all possible measurement outcomes" can be made. If your assumption is mistaken for any reason, however, this might yield to mistaken assessments of nonclassicality. This particular aspect has been gaining special attention among the foundations community and poses an interesting direction for future research [36, 42, 61, 62].

One important feature of simplex embedding is that it will always become possible under the action of depolarising noise [32, 63, 64]. In particular, we define depolarising noise in a GPT as the linear map $D : V \to V$ such that, given a special state $s_\mu \in \Omega$,

$$D(s_P) = \langle s_P, u\rangle\, s_\mu, \quad \forall s_P \in \Omega. \tag{2.19}$$

In this case, it is possible to see that there is $r \in [0, 1)$ such that the GPT $(\Omega_D, \mathcal{E}, V, \langle\cdot,\cdot\rangle)$ admits of a simplex embedding, where

$$\Omega_D := \{(1-r)s_P + r\,\langle s_P, u\rangle\, s_\mu, \quad \forall s_P \in \Omega\}. \tag{2.20}$$

Naturally, $r = 0$ indicates that the GPT already admits of a simplex embedding, without the need of noise, and therefore the quantity r constitutes a well motivated operational measure of nonclassicality. The connection with the standard notion of *robustness to depolarising noise* comes when the vector s_μ represents the maximally mixed state for a quantum system.

[1] In fact, there are additional constraints in the geometry of the sets Ω^A and $\mathcal{E}^A$, but they will play no role for the purposes of this thesis.

The existence of a simplex embedding for a GPT fragment has been proven to be equivalent to the existence of a noncontextual ontological model for the operational scenario (when there is an operational scenario associated to the fragment) [48], and therefore we can look for scenarios that can provide advantage over classical performances in many protocols by employing this mathematical formalism. In Ref. [Second Paper], we introduce a linear program that takes in finite sets of states and effects, and estimates the amount of depolarising noise r that must act on these states so that a simplex embedding becomes possible. Furthermore, the program can also be modified to estimate the robustness r with respect to other noise models. In Ref. [Third Publication], we modify the program to also consider robustness to dephasing noise, and employ it to investigate how much of a role coherence plays in operational scenarios related to minimum-error state-discrimination protocols [19].

2.4 Summary

In this chapter, I introduced the notion of generalised contextuality, a signature of non-classicality that is intuitively motivated and operationally meaningful. By discarding the excess information (i.e., the contexts) from an operational scenario via quotienting, one is left with a minimalistic version of the operational theory that is still capable of providing the same statistical data as the original one. However, the operational description cannot prescribe an ontology for the experiment under investigation, and for that I introduce ontological models. These probabilistic models rely on bayesian probability theory and boolean logic to assign definite values to the physical properties of the system, and the statistical data emerges from sampling these definite values. Finally, I argue that if an operational scenario is to be explained by an ontological model, then its quotiented version should also be explained by the same model. This assumption of noncontextuality is justified by the leibnizian principle of the identity of indiscernibles, that prescribes that an ontology should describe states of affairs that cannot be distinguished by any experiment as the same theoretical object. Quantum theory, however, does not admit of such a noncontextual ontological model, and this has been pointed as the source of quantum advantage in many protocols.

I procede by introducing the two frameworks to investigate contextuality that will be employed in this thesis. First I present the hypergraph approach, a framework in which quotiented operational scenarios are mapped to hypergraphs whose vertices are weighted by the statistical data of the scenario. Despite its formulation being restricted to quotiented scenarios without the need (or possibility) to accommodate more than one equivalence class of preparations, we can still leverage this framework to assess the noncontextual, quantum or beyond-quantum realisability of a data set. Moreover, the hypergraph approach introduces a hierarchy of semidefinite programs that converges to the characterisation of the quantum realisable correlations for a scenario. Its first level, in fact, is closely related to the notion of macroscopic contextuality employed in Ref. [First Paper], and for Bell scenarios, it coincides with the almost quantum set of correlations.

We then introduce the GPT framework with which Refs. [Second Paper], [Third Paper], [Repository], and [Preprint] were developed. In this framework, the quotiented operational theories are mapped to real vector spaces equipped with an inner product. Preparation and measurement outcome equivalence classes are mapped to vectors in this space (named respectively states and effects), and the statistics are recovered by the inner product. I then discuss some of the additional properties a GPT must have, such as that the sets of states and effects must be tomographic to one another and proceed to relax these requirements so that operational scenarios that are not full theories can also be described by the framework, defining GPT fragments and their accessible counterparts. Finally, I introduce the notion of simplex embedding of a GPT, that has been proven to be equivalent to the existence of a noncontextual ontological model for the associated operational scenario, and define an operational measure for the simplex-embeddability of a GPT as the amount of partial noise it must receive to admit of such an embedding, given a particular noise model. This quantity is crucial for Refs. [Second Paper], [Third Paper], and [Preprint], that respectively introduce a linear program for computing this quantity and extend it for other types of noise in order to investigate specific communication tasks.

In the coming chapter, I summarise for each paper in this thesis the context in which they were developed, the ideas and tools employed by them and the results of our investigations.

Chapter 3

Summary of dissertation

3.1 *On characterising assemblages in Einstein-Podolsky-Rosen scenarios* [First Paper]

I begin by summarising our investigations on to what extent noncontextuality can be leveraged as a notion of classicality to characterise assemblages in Einstein-Podolsky-Rosen (EPR) scenarios. In this work, we make a formal summary of the EPR formalism, focusing especially on tools for charactersing sets of assemblages with respect to their possible realisations within classical, quantum, and beyond-quantum theories. Later, we argue that EPR scenarios are ultimately prepare-and-measure, and thus we can explore the role of *macroscopic noncontextuality*, a theoretical principle for correlations, in characterising these assemblages. We conclude that all almost-quantum assemblages must satisfy this assumption, but they are not fully characterised by it. Although not directly connected to the essence of this thesis, this publication further motivates the need for device-dependent notions of nonclassicality in order to characterise quantum advantage in many tasks, with generalised contextuality as our primary example.

3.1.1 EPR scenarios and realisability

Einstein-Podolsky-Rosen scenarios are an example of a causal structure in which distant parties receive shares of a prepared system, in a similar spirit as Bell scenarios [6] and entanglement scenarios [8]. These causal structures are coined *common-cause*, and what changes from one to the other is which parties can characterise their systems as being quantum [1, 2, 43]. In entanglement scenarios, all agents are assumed to know that their shares constitute quantum systems; in Bell scenarios, no parties do so. EPR scenarios are a middle ground between these two extreme cases, in the sense that only one party is assumed to be aware that their share of the common cause is quantum[1].

Differently from Bell scenarios, that are described by the correlations $\{p(ab|xy)\}$ emerging from the measurements performed by the parties, or from entanglement scenarios, that are described by the state ρ shared by the parties, EPR scenarios are described by *assemblages*: a collection of subnormalised states satisfying some constraints. In particular, consider that all uncharacterised parties A_i, $i = 1, ..., n$ perform measurements $x \in \mathbb{X}$ with outcome labels $a \in \mathbb{A}$, and that the characterised party B describes their share of the system with a finite-dimensional Hilbert space $\mathcal{H}_B$. An assemblage is given by the set

[1] In fact, there can be more than one characterised party in an EPR scenario, although this is indistinguishable from the case with a single characterised party for the purposes of EPR nonclassicality.

$\Sigma_{\mathbb{A}|\mathbb{X}} := \{\sigma_{\vec{a}|\vec{x}}\}_{\vec{a}\in\mathbb{A}^{\times n},\vec{x}\in\mathbb{X}^{\times n}}$, such that $\sigma_{\vec{a}|\vec{x}} \in \mathcal{D}(\mathcal{H}_B)$ for all $\vec{a}$ and $\vec{x}$, and such that

$$\sum_{a_i} \sigma_{a_1,\dots,a_i,\dots,a_n|x_1,\dots,x_i,\dots,x_n} = \sum_{a_i} \sigma_{a_1,\dots,a_i,\dots,a_n|x_1,\dots,x_i',\dots,x_n}, \quad \forall x_i, x_i' \in \mathbb{X}, \forall a_i \in \mathbb{A}, i = 1, \dots, n; \tag{3.1}$$

$$\mathrm{tr}\left(\sum_{\vec{a}} \sigma_{\vec{a}|\vec{x}}\right) = 1, \quad \forall \vec{x} \in \mathbb{X}^{\times n}. \tag{3.2}$$

that is, ignoring the outcome obtained by any party implies having no information on their measurement choice. This assumption is named *no signalling*, and the set of assemblages satisfying it is what we will explore. In particular, an assemblage is *quantum* when there is a finite-dimensional, factorised Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, a state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and projective measurements $\{\Pi_{a_i|x_i}\}_{a_i\in\mathbb{A},x_i\in\mathbb{X},i=1,\dots,n}$ over $\mathcal{H}_A$, such that

$$\sigma_{\vec{a}|\vec{x}} = \mathrm{tr}_A\left((\Pi_{a_1|x_1}\dots\Pi_{a_n|x_n}) \otimes \mathbb{1}\,|\Psi\rangle\langle\Psi|\right), \quad \forall \vec{a} \in \mathbb{A}^{\times n}, \vec{x} \in \mathbb{X}^{\times n}, \tag{3.3}$$

and such that

$$i \neq j \implies [\Pi_{a_i|x_i}, \Pi_{a_j|x_j}] = 0, \quad \forall i, j = 1, \dots, n,\ a_i, a_j \in \mathbb{A},\ x_i, x_j \in \mathbb{X}. \tag{3.4}$$

However, not all no-signalling assemblages have quantum realisation. In fact, it has been demonstrated that beyond-quantum realisations are possible for EPR scenarios with $n = 2$ [65]. In particular, there are sets of assemblages that include but are not restricted to quantumly-realisable ones. One set that is of particular interest is the set of *almost-quantum assemblages*, that can be defined as a relaxation of the quantum set. Similarly, an assemblage has an almost-quantum realisation when there is a finite-dimensional, factorised Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$, a state $|\Psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$ and projective measurements $\{\Pi_{a_i|x_i}\}_{a_i\in\mathbb{A},x_i\in\mathbb{X},i=1,\dots,n}$ over $\mathcal{H}_A$, such that

$$\sigma_{\vec{a}|\vec{x}} = \mathrm{tr}_A\left((\Pi_{a_1|x_1}\dots\Pi_{a_n|x_n}) \otimes \mathbb{1}\,|\Psi\rangle\langle\Psi|\right), \quad \forall \vec{a} \in \mathbb{A}^{\times n}, \vec{x} \in \mathbb{X}^{\times n}, \tag{3.5}$$

and such that

$$\Pi_{a_1|x_1}\dots\Pi_{a_n|x_n} \otimes \mathbb{1}\,|\Psi\rangle = \Pi_{a_{\pi(1)}}\dots\Pi_{a_{\pi(n)}} \otimes \mathbb{1}\,|\Psi\rangle, \quad \forall \vec{a} \in \mathbb{A}^{\times n}, \vec{x} \in \mathbb{X}^{\times n},\ \forall \pi \in \mathsf{Sym}(n), \tag{3.6}$$

where π is a permutation. Therefore, we do not require any longer that the projectors for each of the uncharacterised parties A_i commute between the parties, but only that they can be permutated for this particular quantum state that realises the assemblage. It has been demonstrated that this set is strictly larger than the set of quantum assemblages [65], and that alternatively one can characterise an assemblage as almost-quantumly realisable as an instance of a semidefinite program. In Ref. [First Paper], we give a formal definition of this method.

It is known that for the case of a single uncharacterised party ($n = 1$), all non-signalling assemblages have a quantum realisation. This result has been proven some times in literature by different sources [66, 67], but we provide a novel derivation of it in terms of almost quantum certificates.

3.1.2 EPR assemblages as correlations

A reader that is familiarised with the characterisation of Bell correlations will easily recognise the motivation of each of the sets of assemblages mentioned previously. Indeed, the vast majority of literature on characterising the classical or quantum realisability of multiparty scenarios, especially when it comes to physical principles, has been developed to correlations rather than assemblages [51–56]. The core of this publication lies on the argument that EPR assemblages are ultimately correlations in prepare-and-measure scenarios (more specifically, in Bell scenarios). Indeed, telling that the party B has a characterised system described by a finite-dimensional Hilbert space $\mathcal{H}_B$ boils down to saying that this party *knows* that this is a system theoretically described by this Hilbert space, and moreover has access to a *tomographically complete* set of measurements $\mathrm{TC}_B := \{M_{b|y}\}_{b\in\mathbb{B},y\in\mathbb{Y}}$. We therefore introduce the sets of the tomographic data from party B, defined as

$$\mathbb{P}_{\mathbb{AB}|\mathbb{XY}} := \{\mathrm{tr}(M_{b|y}\sigma_{\vec{a}|\vec{x}}) : \sigma_{\vec{a}|\vec{x}} \in \Sigma_{\mathbb{A}|\mathbb{X}}, M_{b|y} \in \mathrm{TC}_B\}, \tag{3.7}$$

to make statements about assemblages. More specifically, we will say that an assemblage $\Sigma_{\mathbb{A}|\mathbb{X}}$ satisfies some physical principle if its tomographic data $\mathbb{P}_{\mathbb{AB}|\mathbb{XY}}$ satisfies that physical principle. This approach is advantageous since we can quickly import correlation-based physical principles to the context of EPR scenarios.

Notice, for instance, that every almost-quantum assemblage comes from an almost-quantum set of tomographic data [39]. To see that, consider $\Sigma_{\mathbb{A}|\mathbb{X}}$ an almost-quantum assemblage, i.e., there are $\mathcal{H}_A\otimes\mathcal{H}_B$, $|\Psi\rangle \in \mathcal{H}_A\otimes\mathcal{H}_B$ and $\{\Pi_{a_i|x_i}\}_{a_i\in\mathbb{A},x_i\in\mathbb{X},i=1,\dots,n}$ satisfying Eq. 3.5 and 3.6. The tomographic data $\mathbb{P}_{\mathbb{AB}|\mathbb{XY}}$ characterising this assemblage has a realisation as follows:

- There is a finite-dimensional Hilbert space $\mathcal{H} := \mathcal{H}_A \otimes \mathcal{H}_B$, and a set $|\Psi\rangle \in \mathcal{H}$;
- There are projective measurements $\{\tilde{\Pi}_{a_i|x_i} := \Pi_{a_i|x_i} \otimes \mathbb{1}\}_{a_i\in\mathbb{A},x_i\in\mathbb{X},i=1,\dots,n} \cup \{\tilde{M}_{b|y} := \mathbb{1} \otimes M_{b|y}\}_{b\in\mathbb{B},y\in\mathbb{Y}}$ acting on $\mathcal{H}$. In particular, $\{M_{b|y}\}_{b\in\mathbb{B},y\in\mathbb{Y}}$ are a tomographic complete set for the Hilbert space $\mathcal{H}_B$.
- The projective measurements are such that, for all $\vec{a} \in \mathbb{A}^{\times n}, \vec{x} \in \mathbb{X}^{\times n}, b \in \mathbb{B}, y \in \mathbb{Y}$,

$$\tilde{\Pi}_{a_1|x_1}...\tilde{\Pi}_{a_n|x_n}\tilde{M}_{b|y}\,|\Psi\rangle = \pi(\tilde{\Pi}_{a_1|x_1}...\tilde{\Pi}_{a_n|x_n}\tilde{M}_{b|y})\,|\Psi\rangle\,, \tag{3.8}$$

  where π permutates the ordering of the product. This holds since $\tilde{M}_{b|y}$ commutes with any $\tilde{\Pi}_{a_i|x_i}$.
- Correlations in $\mathbb{P}_{\mathbb{AB}|\mathbb{XY}}$ are given by

$$p(\vec{a}b|\vec{x}y) = \langle\Psi|\tilde{\Pi}_{a_1|x_1}...\tilde{\Pi}_{a_n|x_n}\tilde{M}_{b|y}|\Psi\rangle\,. \tag{3.9}$$

Since this is one of the definitions of an almost-quantum set of correlations, we can conclude that $\mathbb{P}_{\mathbb{AB}|\mathbb{XY}}$ has an almost-quantum realisation.

3.1.3 Macroscopic assemblages and macroscopic non-contextuality

We will now apply our reasoning by importing a physical principle for correlations known as *macroscopic noncontextuality.* This principle, motivated by the notion of *macroscopic locality* [52], defines for every prepare-and-measure scenario with a single equivalence class of preparations a macroscopic version of this scenario. This version consists of N non-interacting copies of the original system, all prepared in the same equivalence class, and such that all copies undergo the same measurement $x \in \mathbb{X}, y \in \mathbb{Y}$, triggering different outcomes $a \in \mathbb{A}, b \in \mathbb{B}$ for each copy. The detectors are moreover assumed to have a bounded resolution, such that on average $\sqrt{N}$ clicks are required for a detector k to click. When the limit of $N \to \infty$ is taken, this yields to a Gaussian distribution for the intensities

$I_{k|M}$ on each detector. The macroscopic version of this scenario is characterised by the distribution $p_{macro}(\vec{I}_{x_1}, \vec{I}_{x_2}, ..., \vec{I}_y)$.

By employing the hypegraph approach [49] (see Sec. 2.2), we can say that such a distribution will admit of a noncontextual ontological model when there is an ontic space given by all the intensity vectors for all possible measurements $x_i \in \mathbb{X}$ and $y \in \mathbb{Y}$, that is,

$$\lambda := (\vec{I}_{x_1=1}, ...\vec{I}_{x_1=|\mathbb{X}|}, \vec{I}_{x_2=1}, ..., \vec{I}_{x_2=|\mathbb{X}|}..., \vec{I}_{y=1}, ...\vec{I}_{y=|\mathbb{Y}|}), \tag{3.10}$$

a probability distribution

$$p_{NC}(\lambda) := p(\vec{I}_{x_1=1}, ...\vec{I}_{x_1=|\mathbb{X}|}, \vec{I}_{x_2=1}, ..., \vec{I}_{x_2=|\mathbb{X}|}..., \vec{I}_{y=1}, ...\vec{I}_{y=|\mathbb{Y}|}), \tag{3.11}$$

such that the macroscopic probabilities are given by marginalising over the measurements that were not implemented,

$$p_{macro}(\vec{I}_{x_1}, \vec{I}_{x_2}, ..., \vec{I}_y) = \int p_{NC}(\lambda) \prod_{x_i' \neq x_i} d\vec{I}_{x_i'} \prod_{y' \neq y} d\vec{I}_{y'}. \tag{3.12}$$

Ref. [55] shows that this assumption will be satisfied if and only if the underlying statistics $\{p(\vec{a}b|\vec{x}y)\}_{\vec{a}\in\mathbb{A}^{\times n}, \vec{x}\in\mathbb{X}^{\times n}, b\in\mathbb{B}, y\in\mathbb{Y}}$ are $\mathcal{Q}_1$ (see Sec. 2.2). What is most important is that, when the scenario is a Bell scenario, the set $\mathcal{Q}_1$ is known to be equivalent to the set of almost quantum correlations.

This, combined with our reasoning from the previous section, allows us to conclude that all almost quantum assemblages satisfy the assumption of macroscopic noncontextuality, since as we argued, any tomographic data obtained from almost quantum assemblages is almost-quantumly realisable. Since any data set obtained by an EPR scenario is compatible with a Bell scenario, we can tell that this means that almost quantum assemblages are macroscopic noncontextual.

The converse, however, is not true: there are sets of almost quantum correlations that do not constitute a set of tomographic data for an assemblage. There are for instance examples of almost-quantum assemblages that do not have a quantum realisation, but that any statistics obtained from them are quantumly realisable [65]. This means that any physical principle that bounds the set of quantum correlations without making any reference to the existence of a characterised party will not be characterising the set of quantum assemblages. In Ref. [First Paper] we provide a stronger definition of macroscopic noncontextuality that includes device-dependent assumptions such as the existence of B's Hilbert space and tomographic measurements. Once this is assumed, we can demonstrate that this subset of almost quantum correlations implies the existence of almost quantum assemblages.

This fact builds up on previous evidence [36, 68] to point that characterising nonclassical and even beyon-quantum phenomena eventually requires assumptions about the underlying structure of the system. This invites us to shift our approach to contextuality to the GPT framework. In the following Section, I summarise the contribution of Ref. [Second Paper] to characterising generalised contextuality by introducing a linear program that assesses the existence of a simplex embedding for prepare and measure scenarios.

3.2 *Linear program for testing nonclassicality and an open-source implementation* [Second Paper]

As described in Chapter 2, checking for the existence of a noncontextual ontological model for a particular operational scenario might not be straightforward, but one can adopt the GPT framework and leverage its mathematical features to instead check for the existence of a simplex embedding for the associated GPT fragment. In this work, we show that checking for this embedding is a linear program, and provide an implementation in Mathematica that executes this task [69]. Moreover, we obtain an upper bound on the dimension of the ontic space of a simplex-embeddable accessible fragment. It still remains an open question whether this bound is tight or not. We then apply this implementation to simple examples of prepare-and-measure scenarios that belong to the stabiliser subtheory [70], quantum theory, and boxworld [71]. In this Section, I summarise these results using standard algebraic notation instead of the diagrammatic representation employed in the publication, so that readers that are unfamiliar with diagrams have an alternative ground to assess our results.

3.2.1 Positive-cone facets

As introduced in Chapter 2, an accessible GPT fragment is defined as the sets Ω^A and $\mathcal{E}^A$ of states and effects described in the subspaces they span, along with the inclusion maps I_Ω and $I_\mathcal{E}$ that recover the inner product. We can moreover characterise the *positive cones* $\mathsf{Cone}(\Omega^A)$ and $\mathsf{Cone}(\mathcal{E}^A)$ as the sets of states and measurements that yield non-negative values via the inner product. Formally, these are defined as

$$\mathsf{Cone}(\Omega^A) := \{s = \sum_i q_i s_i, \quad \forall q_i \in \mathbb{R}^{\geq 0}, \forall s_i \in \Omega^A\}; \tag{3.13}$$

$$\mathsf{Cone}(\mathcal{E}^A) := \{e = \sum r_i e_i, \quad \forall r_i \in \mathbb{R}^{\geq 0}, \forall e_i \in \mathcal{E}^A\}. \tag{3.14}$$

In general, positive cones are a useful tool for investigating properties of GPTs, since they underpin an ordering for the sets of states and effects [72, 73]. Physically, the positive cone of states or effects contains all the *logically possible* states or effects for the accessible fragment, and therefore the notion of *cone equivalence*, i.e., when two accessible fragments share the same positive cones of states and effects, subsumes the idea that both scenarios share the same logically possible (equivalence class of) preparations and measurement outcomes [60]. This notion has been employed to build a class of scenarios that are not simplex-embeddable and nevertheless do not display any incompatibility of measurements, disproving the standard belief that incompatibility is necessary for contextuality [74]. It has also been demonstrated that if an accessible GPT fragment admits of a simplex embedding, all cone-equivalent fragments will admit of a simplex embedding as well [60].

Alternatively, one can define $\mathsf{Cone}(\Omega^A)$ and $\mathsf{Cone}(\mathcal{E}^A)$ by characterising its facet inequalities. These are finite sets of vectors[2] $\{h_i^\Omega\}_{i=1}^n$ and $\{h_j^\mathcal{E}\}_{j=1}^m$ such that

$$\langle I_\Omega(v), I_\mathcal{E}(h_i^\Omega)\rangle \geq 0 \iff v \in \mathsf{Cone}(\Omega^A), \quad \forall v \in \mathsf{LinSpan}(\Omega^A); \tag{3.15}$$

$$\langle I_\Omega(h_j^\mathcal{E}), I_\mathcal{E}(w)\rangle \geq 0 \iff w \in \mathsf{Cone}(\mathcal{E}^A), \quad \forall w \in \mathsf{LinSpan}(\mathcal{E}^A). \tag{3.16}$$

For instance, in the case of quantum theory, the vectors v, w, h_i^Ω and $h_j^\mathcal{E}$ are represented by Hermitian operators acting over the Hilbert space $\mathcal{H}$, and the inner product is captured by the trace of the product between the two corresponding operators.

One can therefore collect the facet vectors $\{h_i^\Omega\}$ and $\{h_j^\mathcal{E}\}$ in the arrays H_Ω and $H_\mathcal{E}$, such that

$$H_\Omega(v) := \begin{pmatrix} \langle I_\Omega(v), I_\mathcal{E}(h_1^\Omega)\rangle \\ \vdots \\ \langle I_\Omega(v), I_\mathcal{E}(h_n^\Omega)\rangle \end{pmatrix}, \quad \forall v \in \mathsf{LinSpan}(\Omega^A); \tag{3.17}$$

[2] The fact that they are finite follows from the fact that Ω^A and $\mathcal{E}^A$ are themselves finite, and from McMullen's upper bound theorem [75].

$$H_{\mathcal{E}}(w) := \begin{pmatrix} \langle I_\Omega(h_1^{\mathcal{E}}), I_{\mathcal{E}}(w)\rangle \\ \vdots \\ \langle I_\Omega(h_m^{\mathcal{E}}), I_{\mathcal{E}}(w)\rangle \end{pmatrix}, \quad \forall w \in \mathsf{LinSpan}(\mathcal{E}^A). \tag{3.18}$$

Notice that, with this construction, we can rewrite the definition of the positive cone, since

$$v \in \mathsf{Cone}(\Omega^A) \iff H_\Omega(v) \geq_e 0, \tag{3.19}$$

and

$$w \in \mathsf{Cone}(\mathcal{E}^A) \iff H_{\mathcal{E}}(w) \geq_e 0, \tag{3.20}$$

where $\geq_e$ stands for entry-wise non-negativity.

3.2.2 Linear program for simplicial cone embedding

Numerically, we represent the sets Ω^A, $\mathcal{E}^A$ as arrays of vectors, i.e., real matrices V_Ω with $\mathsf{dim}(\Omega^A)$ rows and $|\Omega^A|$ columns for the states, and $V_{\mathcal{E}}$ with $\mathsf{dim}(\mathcal{E}^A)$ rows and $|\mathcal{E}^A|$ columns for the effects. In this representation, the inclusion maps I_Ω and $I_{\mathcal{E}}$ become real matrices $\mathsf{dim}(V) \times \mathsf{dim}(\Omega^A)$ and $\mathsf{dim}(V) \times \mathsf{dim}(\mathcal{E}^A)$, respectively, where V is again the real vector space in which the GPT fragment associated to this accessible fragment lives in.

Characterising the facets of the positive cones can be done by reverting the Double Description Method [76], a method for generating a list of extremal vertices and edges of a convex polytope given a collection of linear inequalities. This is done via the cdd library once the sets of states and effects in the GPT fragment are provided.

Once the positive-cone facets H_Ω and $H_{\mathcal{E}}$ and the inclusion maps I_Ω and $I_{\mathcal{E}}$ of the accessible GPT fragment are known, one can find whether there is a simplex-embedding for the scenario by instead asking whether there is a substochastic map $\sigma : \mathbb{R}^n \to \mathbb{R}^m$, i.e., a matrix $\sigma \geq_e 0$ such that

$$I_{\mathcal{E}}^T \cdot I_\Omega = H_{\mathcal{E}}^T \cdot \sigma \cdot H_\Omega, \tag{3.21}$$

where $\cdot$ is the standard matrix product. The reason why this is equivalent to the existence of a simplex embedding is that, if there is a simplex $\Delta \subset \mathbb{R}^\Lambda$ and linear maps ι, κ as defined in Chapter 2.3.2, then

$$\iota(s) = v_s \in \Delta, \quad \forall s \in \Omega^A. \tag{3.22}$$

Because this map is linear, it is also the case that

$$\iota(v) \in \mathsf{Cone}(\Delta), \quad \forall v \in \mathsf{Cone}(\Omega^A). \tag{3.23}$$

This means that for an orthonormal basis $\{\vec{\lambda}\}$ of $\mathbb{R}^\Lambda$, given that we can represent ι as a matrix $\Lambda \times \mathsf{dim}(\Omega^A)$, then the covector $\vec{\lambda}^T \cdot \iota$ constitutes a linear combination of the facet vectors $\{h_i^\Omega\}_{i=1}^n$, i.e.,

$$\vec{\lambda}^T \cdot \iota = \sum_i \alpha_i h_i^\Omega, \quad \alpha_i \geq 0, \forall i = 1, ..., n. \tag{3.24}$$

Notice that this does capture Eq. 3.23, since for all $\vec{\lambda}$ in the orthonormal basis, $v \in \mathsf{Cone}(\Omega^A) \implies \vec{\lambda}^T \cdot \iota \cdot v \geq 0$. We can therefore write

$$\iota = \alpha \cdot H_\Omega, \tag{3.25}$$

where $\alpha \geq_e 0$ is a $\Lambda \times n$ matrix. The reasoning follows similarly for κ, so that we can always write

$$\kappa = \beta \cdot H_{\mathcal{E}}, \tag{3.26}$$

with $\beta \geq_e 0$ a $\Lambda \times m$ matrix. Since the existence of a simplex embedding also demands that the matrices ι and κ satisfy

$$I_{\mathcal{E}}^T \cdot I_\Omega = \kappa^T \cdot \iota, \tag{3.27}$$

we can use the above results to rewrite it as

$$I_{\mathcal{E}}^T \cdot I_\Omega = H_{\mathcal{E}}^T \cdot \beta^T \cdot \alpha \cdot H_\Omega, \tag{3.28}$$

and defining $\sigma := \beta^T \cdot \alpha$, we return to Eq. 3.21, since $\sigma \geq_e 0$ by construction. The reverse path is also true: given that there is σ satisfying Eq. 3.21, we can also define the embedding maps ι and κ such that $\iota := H_\Omega$ and $\kappa := H_{\mathcal{E}} \cdot \sigma$, consolidating the equivalence between Eq. 3.21 and the existence of a simplex embedding for the accessible fragment.

I also mention in Chapter 2.3.2 that one important aspect of simplex-embeddability is that there is always a finite amount of partial depolarising noise under which an accessible GPT fragment will admit of a simplex embedding. In this case, let $D : V \to V$ be the depolarising map acting on the full GPT from which the accessible fragment is derived. If we define the set of partially depolarised accessible states as

$$\Omega_D^A(r) := \{s_D := (1-r)s + rP_\Omega \circ D \circ I_\Omega(s), \quad \forall s \in \Omega^A\}, \tag{3.29}$$

the probability rule for all states $s_D \in \Omega_D^A(r)$ and effects $e \in \mathcal{E}^A$ in the depolarised accessible fragment will be given by

$$\langle I_\Omega(s_D), I_{\mathcal{E}}(e)\rangle \,. \tag{3.30}$$

By replacing s_D with its explicit definition, and by adopting the matrix notation introduced earlier, we have that a simplex embedding will exist whenever there is $0 \leq r \leq 1$ and $\sigma \geq_e 0$ such that

$$(1-r)I_{\mathcal{E}}^T \cdot I_\Omega + rI_{\mathcal{E}}^T \cdot D \cdot I_\Omega = H_{\mathcal{E}}^T \cdot \sigma \cdot H_\Omega. \tag{3.31}$$

This in fact constitutes the Linear Program of this paper:

Linear Program. *Let $(\Omega^A, \mathcal{E}^A, I_\Omega, I_{\mathcal{E}})$ be an accessible GPT fragment, and let $(H_\Omega, H_{\mathcal{E}})$ be the positive-cone facets for the sets of states and effects in this accessible fragment. If r is the minimum partial depolarising noise needed for the accessible fragment to admit of a simplex embedding, then it can be computed by the program:*

$$\begin{aligned} &\text{minimize} && r && (3.32)\\ &\text{such that} && (1-r)I_{\mathcal{E}}^T \cdot I_\Omega + rI_{\mathcal{E}}^T \cdot D \cdot I_\Omega = H_{\mathcal{E}}^T \cdot \sigma \cdot H_\Omega, && (3.33)\\ & && \sigma \geq_e 0 \text{ is an } m \times n \text{ matrix.} && (3.34) \end{aligned}$$

3.2.3 Bound on the dimension of the simplicial space

Finally, we provide a bound on the maximum number of vertices a simplex must have for a simplex embedding to exist given an accessible GPT fragment. For that, let us assume that a particular accessible fragment $(\Omega^A, \mathcal{E}^A, I_\Omega, I_{\mathcal{E}})$ admits of a simplex embedding. This means that

$$I_{\mathcal{E}}^T \cdot I_\Omega + H_{\mathcal{E}}^T \cdot \sigma \cdot H_\Omega, \quad \sigma \geq_e 0, \tag{3.35}$$

where H_Ω and $H_{\mathcal{E}}$ are the arrays of facet inequalities introduced previously. We can therefore understand the probability rule $I_{\mathcal{E}}^T \cdot I_\Omega$ as a process belonging to a positive cone $\mathcal{C}$, i.e., the set of $\mathsf{dim}(\mathcal{E}^A) \times \mathsf{dim}(\Omega^A)$ real matrices L such that

$$L := \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_{ij} (h_j^{\mathcal{E}})^T \cdot h_i^\Omega, \quad \gamma_{ij} \geq 0, \forall i,j. \tag{3.36}$$

It follows from Carathéodory's theorem that any matrix living in this cone can be written as a conical combination of no more than $\mathsf{dim}(\mathcal{E}^A)\cdot\mathsf{dim}(\Omega^A)$ vertices in $\mathcal{C}$. This means that there are up to $\mathsf{dim}(\mathcal{E}^A)\cdot\mathsf{dim}(\Omega^A)$ non-zero coefficients χ_{ij} such that

$$I_{\mathcal{E}}^T \cdot I_\Omega = \sum_{i=1}^{n}\sum_{j=1}^{m} \chi_{ij}(h_j^{\mathcal{E}})^T \cdot h_i^\Omega. \tag{3.37}$$

We can now just relabel the summation such that $\xi_{ij} \mapsto \tilde{\xi}_k$, with $k = 1, ..., \mathsf{dim}(\mathcal{E}^A)\cdot\mathsf{dim}(\Omega^A)$. The facet inequalities are relabelled such that $h_i^\Omega \mapsto h_{a(k)}^\Omega$ and $h_j^{\mathcal{E}} \mapsto h_{b(k)}^{\mathcal{E}}$, with $a(k) = i$ and $b(k) = j$, $\forall k \equiv ij$. We thus have

$$I_{\mathcal{E}}^T \cdot I_\Omega = \sum_{k=1}^{\mathsf{dim}(\mathcal{E}^A)\cdot\mathsf{dim}(\Omega^A)} \bar{\xi}_k (h_{b(k)}^{\mathcal{E}})^T \cdot h_{a(k)}^\Omega. \tag{3.38}$$

We can therefore define a square matrix $\tilde{\xi}$ with dimension $\mathsf{dim}(\mathcal{E}^A)\cdot\mathsf{dim}(\Omega^A)$, diagonal with no null elements in its diagonal, such that

$$I_{\mathcal{E}}^T \cdot I_\Omega = \tilde{H}_{\mathcal{E}}^T \cdot \tilde{\xi} \cdot \tilde{H}_\Omega, \tag{3.39}$$

where $\tilde{H}_{\mathcal{E}}$ and $\tilde{H}_\Omega$ are the relabelled arrays of facets. This provides an upper bound to the maximal dimension of the simplex in which an embedding can exist. In the particular case of GPTs, we have $\mathsf{dim}(\Omega) = \mathsf{dim}(\mathcal{E}) = \mathsf{dim}(V)$, so this upper bound reduces to $(\mathsf{dim}(V))^2$. It is however unclear for which circumstances this bound is tight, and how to efficiently derive it from the linear program.

3.2.4 Examples

As example applications of the program, consider the scenario in which a qubit can be prepared in 4 different states, represented by

$$\Omega^F := \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \right\}. \tag{3.40}$$

Consider moreover that there are two possible measurements to be implemented, given by these very same Pauli opereators. This means that the set of effects in the fragment is given by

$$\mathcal{E}^F := \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}. \tag{3.41}$$

In this case, the linear program finds that there is a matrix σ for $r = 0$, which means that no depolarising noise is needed for a simplex embedding to exist. Indeed, this is an example in which all states and effects belong to the stabiliser subtheory, that is known to be noncontextual [44].

Consider now that we rotate the effects along the Y axis, so that all of them form an angle $\theta = \frac{\pi}{4}$ with respect to all states. This means that the set of states Ω^F remains unchanged, while the set of effects is now given by

$$\begin{aligned} \mathcal{E}^F \; := \; & \left\{ \frac{1}{4}\begin{pmatrix} 2+\sqrt{2} & \sqrt{2} \\ \sqrt{2} & 2-\sqrt{2} \end{pmatrix}, \frac{1}{4}\begin{pmatrix} 2-\sqrt{2} & \sqrt{2} \\ \sqrt{2} & 2+\sqrt{2} \end{pmatrix}, \frac{1}{4}\begin{pmatrix} 2-\sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 2+\sqrt{2} \end{pmatrix}, \right. \\ & \left. \frac{1}{4}\begin{pmatrix} 2+\sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 2-\sqrt{2} \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right\}. \end{aligned} \tag{3.42}$$

In this scenario, the code states the impossibility of a simplex embedding in the absence of noise, and provides that the minimum amount of partial depolarising noise required for such an embedding to exist is equal to $r = 1 - \frac{1}{\sqrt{2}}$. For both examples above, the code also provides arrays of epistemic states and response functions corresponding to the noncontextual ontological model of the depolarised scenario. They are explicitly described in Ref. [Second Paper], and we moreover provide additional examples of usage of the program for assessing the nonclassicality in quantum scenarios beyond the qubit, as well as in theories beyond quantum.

In the next Section, I comment on how Ref. [Third Paper] showcases an application of the linear program that goes beyond merely stating whether or not a particular prepare-and-measure scenario satisfies the assumption of noncontextuality. The results of this paper have been verified experimentally [77], and the linear program modification introduced there has also been employed to establish a relation between different resource quantifiers in parity-oblivious multiplexing tasks in Ref. [Preprint], as well as to assess the nonclassicality in an experimental implementation of the quantum interrogation task [78].

3.3 *Contextuality with vanishing coherence and maximal robustness to dephasing* [Third paper]

This publication is an immediate application of the linear program we developed in Ref. [Second Paper] summarised in the previous Section, and aims to explore the relation between the existence of quantum coherence and the observation of contextuality in prepare-and-measure scenarios. In this work, we introduce a family of prepare-and-measure scenarios realised on a qubit that are parametrised by the amount of coherence with respect to the Z axis of the Bloch sphere. We then employ the linear program to numerically investigate the relation between robustness of contextuality to depolarisation and the amount of coherence in the preparations and measurements, and conclude that contextuality exists even when coherence is vanishing. Finally, we modify the linear program to estimate robustness of contextuality to dephasing noise with respect to a given basis, and we identify a family of scenarios that constitute proofs of contextuality maximally robust to dephasing noise, even when coherence in the preparations and measurements is vanishing.

3.3.1 Motivation – contextuality implies coherence

The main goal of Ref. [Third Paper] is to better understand the interplay between contextuality and coherence. The reason why this relation is appealing is that contextuality, when manifested in quantum realisations of prepare-and-measure scenarios, implies the presence of coherence in both preparations and measurements. This fact, despite well-known, is not clearly documented in the literature, and therefore we provide a proof in this publication that is hereby reproduced.

Let us assume that $(\Omega^F := \{\rho_P\}, \mathcal{E}^F := \{E_{k|M}\})$ is a quantum GPT fragment, in which the probability rule is given by the inner product. Let us further assume that the set of states does not show coherence, i.e., there is a basis $\{|i\rangle\langle i|\}_{i=1}^{\mathsf{dim}(\mathcal{H})}$ of $\mathcal{H}$ in which all states ρ_i are diagonal. This means that

$$\rho_P = \sum_{i=1}^{\mathsf{dim}(\mathcal{H})} \langle i|\rho_P|i\rangle\, |i\rangle\langle i| \,, \quad \forall \rho_P \in \Omega^F. \tag{3.43}$$

We can therefore build an ontological model for this scenario by establishing an ontic space

$R^{\mathsf{dim}(\mathcal{H})}$. The epistic states are given by the map

$$\mu_P(i) = \langle i|\rho_P|i\rangle\,, \quad \forall \rho_P \in \Omega^F, \tag{3.44}$$

and the response functions, by the map

$$\xi_{k|M}(i) = \langle i|E_{k|M}|i\rangle\,, \quad \forall E_{k|M} \in \mathcal{E}^F. \tag{3.45}$$

Notice that this indeed reproduces the statistics of the fragment, since

$$\begin{aligned}
p(k|M,P) &= \mathrm{tr}(E_{k|M}\rho_P) &(3.46)\\
&= \sum_{i=1}^{\mathsf{dim}(\mathcal{H})} \langle i|E_{k|M}\rho_P|i\rangle &(3.47)\\
&= \sum_{i,j=1}^{\mathsf{dim}(\mathcal{H})} \langle i|E_{k|M}|j\rangle\langle j|i\rangle\langle j|\rho_P|j\rangle &(3.48)\\
&= \sum_{i=1}^{\mathsf{dim}(\mathcal{H})} \langle i|E_{k|M}|i\rangle\langle i|\rho_P|i\rangle &(3.49)\\
&= \sum_{i=1}^{\mathsf{dim}(\mathcal{H})} \xi_{k|M}(i)\mu_P(i). &(3.50)
\end{aligned}$$

Moreover, this model is noncontextual since it already maps equivalence classes (density operators and POVM elements) to the same epistemic states and response functions, so indiscernible elements in the operational level are explained by the same ontological object.

This proof establishes that some amount of coherence is necessary to display contextuality, but it does not concern how much coherence is enough. It is known for instance that coherence is also present noticeably in noncontextual theories, such as in Spekkens' toy model [44]. This motivates the belief that robustness to dephasing noise might also provide a good operational measure of nonclassicality, since at most total dephasing noise is necessary to allow for a simplex embedding in the sense of Ref. [Second Paper], and opens the question of whether there is a minimum amount of coherence that is necessary to yield a proof of contextuality.

3.3.2 Family of coherent prepare-and-measure scenarios

In order to explore the question posed in the last subsection, we choose a family of prepare-and-measure scenarios that is closely related to the minimum-error state-discrimination protocol [32]. These are given by preparations and measurements over a qubit system such that

$$\begin{aligned}
\Omega^F \;:=\; &\left\{\rho_\psi = \frac{1}{2}\begin{pmatrix} 1+\cos\theta & \sin\theta \\ \sin\theta & 1-\cos\theta \end{pmatrix}, \rho_{\bar{\psi}} = \frac{1}{2}\begin{pmatrix} 1-\cos\theta & -\sin\theta \\ -\sin\theta & 1+\cos\theta \end{pmatrix},\right.\\
&\left.\rho_\phi = \frac{1}{2}\begin{pmatrix} 1+\cos\theta & -\sin\theta \\ -\sin\theta & 1-\cos\theta \end{pmatrix}, \rho_{\bar{\phi}} = \frac{1}{2}\begin{pmatrix} 1-\cos\theta & \sin\theta \\ \sin\theta & 1+\cos\theta \end{pmatrix}\right\},
\end{aligned} \tag{3.51}$$

for $\theta \in [0, \frac{\pi}{2})$. Notice that these preparations satisfy the constraint that

$$\frac{1}{2}(\rho_\psi + \rho_{\bar{\psi}}) = \frac{\mathbb{1}}{2} = \frac{1}{2}(\rho_\phi + \rho_{\bar{\phi}}). \tag{3.52}$$

The set of effects is given by the projections onto these preparations, plus two additional projections comprising the Halstrom measurement [79]. Explicitly, this is given by

$$\begin{aligned}\mathcal{E}^F \quad := \quad & \Bigg\{ E_\psi = \frac{1}{2}\begin{pmatrix} 1+\cos\theta & \sin\theta \\ \sin\theta & 1-\cos\theta \end{pmatrix}, E_{\bar{\psi}} = \frac{1}{2}\begin{pmatrix} 1-\cos\theta & -\sin\theta \\ -\sin\theta & 1+\cos\theta \end{pmatrix}, \\ & E_\phi = \frac{1}{2}\begin{pmatrix} 1+\cos\theta & -\sin\theta \\ -\sin\theta & 1-\cos\theta \end{pmatrix}, E_{\bar{\phi}} = \frac{1}{2}\begin{pmatrix} 1-\cos\theta & \sin\theta \\ \sin\theta & 1+\cos\theta \end{pmatrix}, \\ & E_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \Bigg\}. \end{aligned} \tag{3.53}$$

Notice that the E_ψ and $E_{\bar{\psi}}$ belong to the same measurement, since they sum up to the identity. The same is true for the E_ϕ and $E_{\bar{\phi}}$, as well as E_0 and E_1. The parameter θ in this work is shared by the states and effects, otherwise these effects would not be projecting onto the corresponding preparations.

The reason why we choose this parametrisation of the GPT fragment is that the parameter θ can easily be related to many of the most well-known coherence quantifiers. For instance, if we choose as coherence quantifier the trace distance between the state/effect and its fully dephased version (with respect to the Z axis of the Bloch sphere), we have

$$C(\theta) = \left\| \frac{1}{2}\begin{pmatrix} 1+\cos\theta & \pm\sin\theta \\ \pm\sin\theta & 1-\cos\theta \end{pmatrix} - \frac{1}{2}\begin{pmatrix} 1+\cos\theta & 0 \\ 0 & 1-\cos\theta \end{pmatrix} \right\|_1 \tag{3.54}$$

$$= \left\| \frac{1}{2}\begin{pmatrix} 0 & \pm\sin\theta \\ \pm\sin\theta & 0 \end{pmatrix} \right\|_1 \tag{3.55}$$

$$= \frac{1}{2}\mathrm{tr}\left(\sqrt{\frac{1}{4}\begin{pmatrix} 0 & \pm\sin\theta \\ \pm\sin\theta & 0 \end{pmatrix}^2} \right) \tag{3.56}$$

$$= \frac{1}{4}\mathrm{tr}\begin{pmatrix} |\sin\theta| & 0 \\ 0 & |\sin\theta| \end{pmatrix} \tag{3.57}$$

$$= \frac{1}{2}|\sin\theta|. \tag{3.58}$$

Since quantifiers such as the above are well-established in assessing coherence for qubit systems, we can safely state that increasing the value of θ implies an increase on the coherence for the states and effects (for the range of values θ can assume). We therefore investigate how the robustness of contextuality to noise estimated by the linear program relates to this parameter.

3.3.3 Contextuality with vanishing coherence

As mentioned before, these scenarios are closely related to a state-discrimination task for which contextuality is known to be a resource [32]. In particular, they are particular instances of a larger family of prepare-and-measure scenarios with 4 preparations and 6 measurement outcomes, whose statistics are completely defined by three real parameters, hereby denoted s, c, and ϵ. They relate to the states and effects as per Table 3.1, where we omitted the effects $E_{\bar{\psi}}$, $E_{\bar{\phi}}$ and E_1 since their statistics is given by the data-table combined with normalisation.

We know from Ref. [32] that a noncontextual ontological model for this scenario, satisfying the equivalence relations introduced in the previous subsection, will be possible only when

$$s \leq 1 - \frac{c-\epsilon}{2}. \tag{3.59}$$

	ρ_ψ	$\rho_{\bar{\psi}}$	ρ_ϕ	$\rho_{\bar{\phi}}$
E_ψ	$1-\epsilon$	ϵ	c	$1-c$
E_ϕ	c	$1-c$	$1-\epsilon$	ϵ
E_0	s	$1-s$	$1-s$	s

Table 3.1: Data table for prepare-and-measure scenarios satisfying the equivalence relations introduced previously.

In our family of scenarios, we have

$$s = \cos^2 \frac{\theta}{2}; \quad c = \sin^2 \theta, \quad \epsilon = 0. \tag{3.60}$$

If partial depolarising noise acts on the states of the fragment defined above as per Eq. 2.20, the data extracted from the depolarised scenario is given by what is depicted in Table 3.2.

	$\mathcal{D}(\rho_\psi)$	$\mathcal{D}(\rho_{\bar{\psi}})$	$\mathcal{D}(\rho_\phi)$	$\mathcal{D}(\rho_{\bar{\phi}})$
E_ψ	$1-\frac{r}{2}$	$\frac{r}{2}$	$(1-r)\cos^2\theta + \frac{r}{2}$	$(1-r)\sin^2\theta + \frac{r}{2}$
E_ϕ	$(1-r)\cos^2\theta + \frac{r}{2}$	$(1-r)\sin^2\theta + \frac{r}{2}$	$1-\frac{r}{2}$	$\frac{r}{2}$
E_0	$(1-r)\cos^2\frac{\theta}{2} + \frac{r}{2}$	$(1-r)\sin^2\frac{\theta}{2} + \frac{r}{2}$	$(1-r)\sin^2\frac{\theta}{2} + \frac{r}{2}$	$(1-r)\cos^2\frac{\theta}{2} + \frac{r}{2}$

Table 3.2: Data table for the partially depolarised prepare-and-measure scenario.

By comparing the two data tables, it is easy to identify that

$$s = (1-r)\cos^2\frac{\theta}{2} + \frac{r}{2}; \quad c = (1-r)\sin^2\theta + \frac{r}{2}; \quad \epsilon = \frac{r}{2}. \tag{3.61}$$

If the depolarised scenario admits of a noncontextual ontological model, then Eq. 3.60 is satisfied, leading to a lower bound on the partial noise:

$$s \leq 1 - \frac{c-\epsilon}{2} \iff (1-r)\cos^2\frac{\theta}{2} + \frac{r}{2} \leq 1 - \frac{(1-r)\sin^2\theta}{2} \tag{3.62}$$

$$\iff \frac{1-r}{2}(1+\cos\theta) + \frac{r}{2} \leq 1 - \frac{(1-r)\sin^2\theta}{2} \tag{3.63}$$

$$\iff (1-r)(\sin^2\theta + \cos\theta) \leq 1 \tag{3.64}$$

$$\iff r \geq 1 - \frac{1}{\sin^2\theta + \cos\theta}. \tag{3.65}$$

Therefore, the minimum partial depolarising noise required for a noncontextual ontological model to exist is the one which saturates this bound, that is,

$$r_{depol} = 1 - \frac{1}{\sin^2\theta + \cos\theta}. \tag{3.66}$$

We then notice that for the values of θ considered, $r = 0 \iff \theta = 0$. This means that any amount of coherence in the preparations and measurements is enough to obtain a proof of contextuality, constituting the first result of this paper. This also contrasts with what I claimed to be the expectation in Sec. 3.3.1, that since there are theories completely noncontextual and that can exhibit coherence, there would be a some amount of coherence for which contextuality cannot be observed.

3.3.4 Proof of contextuality maximally robust to dephasing noise

We now proceed to show that not only we can find proofs of contextuality with vanishing coherence, but that we can further find such proofs that additionally endure any amount of partial dephasing noise. Dephasing is a noise model that attacks coherence directly, so one would expect proofs of contextuality to be more vulnerable to this noise, given the arguments from Sec. 3.3.1.

First, we employ the same analytical implementation as the one in Sec. 3.3.3, but replacing the noise model for the following:

$$\mathcal{D}_{deph}(\rho) = (1-r)\rho + r\sum_{i=0}^{1} \langle i|\rho|i\rangle\, |i\rangle\langle i|\,, \quad \forall \rho \in \Omega^F. \tag{3.67}$$

We end up with a data table in the spirit of Table 3.2, and by demanding Eq. 3.60 to hold we get the equation

$$r_{deph} = 1 - \frac{1-\cos\theta}{\sin^2\theta}. \tag{3.68}$$

Notice that once again, $r = 0 \iff \theta = 0$, so the same statement can be produced concerning vanishing coherence. In particular, for vanishing coherence, i.e., $\theta \to 0$ we havee $r \to \frac{1}{2}$, which is considerably higher than the maximum robustness achieved for depolarising noise. We then ask if there is some modification in this family of scenarios that can yield an even larger increase in robustness.

To answer this question, we modify the linear program from Ref. [Second Paper] to estimate the robustness to dephasing in an extended family of scenarios in which the effects E_0 and E_1 are now rotated through the ZX plane of the Bloch sphere. Explicitly, the set of states Ω^F remains the same, while the set of effects is now given by

$$\begin{aligned}
\mathcal{E}^F \quad := \quad & \Bigg\{ E_\psi = \frac{1}{2}\begin{pmatrix} 1+\cos\theta & \sin\theta \\ \sin\theta & 1-\cos\theta \end{pmatrix}, E_{\bar{\psi}} = \frac{1}{2}\begin{pmatrix} 1-\cos\theta & -\sin\theta \\ -\sin\theta & 1+\cos\theta \end{pmatrix}, \\
& E_\phi = \frac{1}{2}\begin{pmatrix} 1+\cos\theta & -\sin\theta \\ -\sin\theta & 1-\cos\theta \end{pmatrix}, E_{\bar{\phi}} = \frac{1}{2}\begin{pmatrix} 1-\cos\theta & \sin\theta \\ \sin\theta & 1+\cos\theta \end{pmatrix}, \\
& E_\alpha = \frac{1}{2}\begin{pmatrix} 1+\cos\alpha & \sin\alpha \\ \sin\alpha & 1-\cos\alpha \end{pmatrix}, E_{\bar{\alpha}} = \frac{1}{2}\begin{pmatrix} 1-\cos\alpha & -\sin\alpha \\ -\sin\alpha & 1+\cos\alpha \end{pmatrix}, \\
& \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \Bigg\}.
\end{aligned} \tag{3.69}$$

Our numerical calculations find that when $\alpha = \frac{\pi}{2}$, that is, when E_α and $E_{\bar{\alpha}}$ align with the X axis, we have $r \to 1$ when $\theta \to 0$. Moreover, notice that if any amount of partial dephasing noise acts on this scenario with $\theta \to 0$ (say, $r = 0.9$), we still have a maximally robust proof of contextuality since no amount of noise lesser than total dephasing will be able to allow for a noncontextual model, and yet this dephased version of the scenario has vanishing coherence on both states and effects.

These results showcase the usefulness of the linear program from Ref. [Second Paper] and shows that the relationship between coherence and the failure of noncontextuality is not straightforward: indeed, there are both examples in which coherence is strongly present in completely noncontextual theories [44], and as demonstrated here, examples in which coherence is vanishing and yet contextuality can still be observed. In the following, I introduce some further work that developed around the linear program from Ref. [Second Paper]. In particular, I comment on the Python implementation of this linear program, as well as another application of this tool that explores when robustness to noise can do better than being an operational measure in order to quantify the resourcefulness provided by contextuality in a quantum communication task.

3.4 *A Python importation of the linear program from Ref. [Second Paper]* [Repository]

We follow up a previous work from Ref. [80]. The main functionality of the code is encapsulated in the `SimplexEmbedding` function found in `simplexEmbedding.py`. This will take a set of states, a set of effects, a unit effect and a maximally mixed state (necessarily in vector form), find the accessible GPT fragment representation for the states and effects with the `DefineAccessibleGPTFragment` function, characterise its cone facets, find the minimal amount of noise r necessary for a simplicial-cone embedding with the function `SimplicialConeEmbedding`, and compute a simplex embedding from the result and the respective sets of embedded states and effects, μ and ξ. The function outputs an array $(\mathtt{r}, \mu, \xi)$.

The following files can be found in the Git repository:

- `preprocessing.py`: contains functions to construct the Gell-Mann orthonormal basis of hermitian operators for a Hilbert space of any dimension, and convert states and inputs from the matricial representation to vector representation. In particular, provides the function `fromListOfMatrixToListOfVectors` necessary to convert inputs from operator form to GPT vector form.

- `math_tools.py`: provides mathematical accessories necessary for the main linear program, such as a function `rref` finding the Reduced Row Echelon Form of a matrix in order to determine the dimension of the space spanned by the sets of states and effects in their accessible GPT fragment representation. The functions characterising the positive cone of states and effects are also specified.

- `simplexEmbedding.py`: provides the main functionality of the repository, and auxiliary functions needed for the main computation.

- `examples.py`: Provides example data for testing the main functions. These are the 4 examples explored in Ref. [Second Paper].

An example of the usage of the code is the following:

```
from simplexEmbedding import SimplexEmbedding
import examples

states, effects, unit, mms = example1()
result = SimplexEmbedding(states, effects, unit, mms)
print("Result: Robustness of contextuality (r) = {result[0]}, Epistemic States (μ) =
    {result[1]}, Response Functions (ξ) = {result[2]}")
```

In this example, the printed output should be an array with a number r, which in the case of example 1 from Ref. [Second Paper] is equal to 0; a list μ of 4-dimensional vectors representing the epistemic states, and a list ξ of 4-dimensional vectors representing the response functions. The `Python` version of this code computes the examples from Ref. [Second Paper] at in about 0.02 s per example, contrasted to a performance of about 1 s per example with the Mathematica version.

3.5 *Robustness of contextuality under different types of noise as quantifiers for parity-oblivious multiplexing tasks* [preprint]

With the Python implementation from Ref. [Repository], we extend the modification in Ref. [Third Paper] to estimate robustness to dephasing noise in a subset of strongly self-dual GPTs. We procede to apply this modification in quantum realisations of parity-oblivious multiplexing tasks, in order to investigate how good robustness is as a quantifier of the quantum advantage in these tasks. We derive an analytical result connecting robustness to depolarisation and the success rate in the task, showing that both quantities are monotonically related. We also show how the linear program can be implemented to derive upper bounds on the number of bits that can be optimally encoded in a particular quantum system. Finally, we explore the particular case of parity-oblivious multiplexing encoding 3 bits in 1 qubit, and show that although robustness to dephasing with respect to a specific basis is not as good a quantifier, a slightly symmetrisation of this quantity over key axes lifts it to the same condition as robustness to depolarisation.

3.5.1 Modification of the program

The implementation of Ref. [Second Paper] [69] makes it easy for modifications of the noise model. In particular, Ref. [Third Paper] modifies it by explicitly typing in the matrix representing dephasing noise in the real hemisphere of the Bloch sphere, parametrised by an angle η. It further modifies the code so that instead of asking for a maximally mixed state, it now asks for the value η that will define the axis in the ZX plane with respect to which dephasing will happen. This modification however is restricted to a very particular family of GPT fragments, and it would be convenient to extend the reach of the code to estimate robustness to dephasing at least within the full Bloch sphere, so that it can cover most of the interesting quantum prepare-and-measure scenarios.

We therefore modify the program in Ref. [Repository] such that instead of taking as input a maximally mixed state, it asks for an array of effects $\{e_i\}_{i=1}^m$. The depolarising map is then replaced by the map

$$D_{deph} := (e_1, ..., e_m) \cdot (e_1, ..., e_m)^T, \tag{3.70}$$

where $(e_1, ..., e_m)$ is a matrix with $\mathsf{dim}(V)$ rows and m columns, V being the vector space in which the fragment lives. Intuitively, what $D(s)$ does over a state s is to estimate for each e_i the probability $\langle s, e_i \rangle$, and then preparing the corresponding state e_i. The final state is a convex mixture for all $i = 1, ..., m$. Notice therefore that only GPTs in which the effects have corresponding states given by transposition admit such a dephasing map, and therefore we constrain ourselves to studying strongly self-dual GPTs [72, 81, 82].

As an example, consider the standard dephasing map with respect to the Z basis in a qubit. Following Sec. 2.3, we have

$$|0\rangle\langle 0| \mapsto \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}; \quad |1\rangle\langle 1| \mapsto \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, \tag{3.71}$$

where the factor $\frac{1}{\sqrt{2}}$ ensures that these vectors multiplied with their transposes are equal to 1, since they are pure states. By Eq. 3.70, the dephasing map is given by

$$D_{deph} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.72}$$

Notice that indeed for any (normalised) qubit state, one has

$$D \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ \langle X \rangle \\ \langle Y \rangle \\ \langle Z \rangle \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 0 \\ \langle Z \rangle \end{pmatrix}, \tag{3.73}$$

which is precisely the action of the total dephasing noise.

3.5.2 Parity-oblivious multiplexing

A celebrated result by Holevo [83] poses a bound over the amount of bits that can be recovered from a number of qubits. In particular, if one has n qubits for some communication task, Holevo's theorem states that no more than $\log n$ bits can be recovered from these qubits, and therefore it can be mistakenly understood as a no-go theorem for quantum advantage in communication tasks. However, there are many examples of quantum advantage in communication, since this advantage can stem from the fact that qubits can *store* more information than classical channels, even though not all of this information can be retrieved [84, 85].

In a parity-oblivious task, a string x of n classical bits is encoded in the preparation of a quantum state P_x. Measurements are constructed such that a binary measurement M_y returns outcome 0 when the y-th entry of the string x is 0, and 1 otherwise. Moreover, the encoding must be done so that no information about the parity of the string can be recovered from a single measurement. In other words, given the parity strings $\{t|t = (t_1, ..., t_n), \quad \sum_i t_i \geq 2\}$, the scenario satisfies for all s, x and y

$$\sum_{x|x\cdot t=0} p(k = x_y|M_y, P_x) = \sum_{x|x\cdot t=1} p(k = x_y|M_y, P_x). \tag{3.74}$$

The resourcefulness of this task is usually quantified by the probability of getting the outcome for measurement y that correctly matches the y-th bit in the string x, for any outcome and string, called success rate. Formally, it is given by

$$s = \frac{1}{2^n n} \sum_{y=1}^{n} \sum_{x} p(k = x_y|M_y, P_x). \tag{3.75}$$

It has been shown that quantum realisations of this protocol can exceed the noncontextual bound of $s_{NC} = \frac{1}{2}\left(1 + \frac{1}{n}\right)$ [25], attesting that contextuality is the source of quantum advantage for these tasks, since an optimal encoding with a quantum realisation has been demonstrated to reach up to $s_Q = \frac{1}{2}\left(1 + \frac{1}{\sqrt{n}}\right)$ [27]. Indeed, Ref. [86] recently proved that the success rate s is indeed a good quantifier of contextuality in a resource theory of simplex-embeddability.

3.5.3 Robustness to depolarising noise in the n-to-1 POM task

In the depolarised version of the experiment, the quantum states associated to the bit strings encoded are given by

$$D_{depol}(\rho_x) = (1-r)\rho_x + r\mu, \quad \forall \rho_x \in \mathcal{P}, \tag{3.76}$$

where μ is the maximally mixed state for the system encoding the strings. The success rate for this depolarised scenario is therefore given by

$$\begin{aligned} s_{depol} &= \frac{(1-r)}{2^n n}\sum_{y=1}^{n}\sum_{x\in[0,1]^{\times n}} p(b=x_y|P_x,M_y) + \frac{r}{2^n n}\sum_{y=1}^{n}\sum_{x\in[0,1]^{\times n}} \mathrm{tr}(E_{b=x_y}\mu) &(3.77)\\ &= (1-r)s + \frac{r}{2n}\sum_{y=1}^{n}\mathrm{tr}(\mu) &(3.78)\\ &= (1-r)s + \frac{r}{2}, &(3.79)\end{aligned}$$

The robustness of contextuality to depolarisation is given by the minimum noise necessary that a depolarised scenario needs to perform as poorly as a noncontextual simulation, and so we request $s_{depol} = s_{NC}$, obtaining the relation

$$r^{depol}_{min} = \frac{s - s_{NC}}{s - \frac{1}{2}}. \tag{3.80}$$

This leads us to our first result: In any n-to-1 parity-oblivious multiplexing scenario, success rate and robustness of contextuality to depolarisation are monotonically proportional resource quantifiers via Eq. 3.80 (i.e., the derivative of r^{depol}_{min} with respect to s is always positive in the range of values that s can assume).

We also leverage this result to obtain a method for singling out the maximum number of bits that can be optimally encoded in a qubit. For such, we employ the linear program from Ref. [Repository] to motivate the argument that $r < \frac{1}{2} =: r_{max}$ for any prepare-and-measure scenario with finite sets of states and effects on a qubit. We can easily do this by parametrising a family of equally distributed states and effects across the Bloch sphere, such that the parameter n is the number of such states and effects, and use the linear program to plot robustness to depolarisation as a function of the number of states/effects, and conclude that this quickly converges to $r = \frac{1}{2}$.

We then use the noncontextual [25] and quantum [27] optimal bounds for success rate to derive the relation

$$n_{max} < \frac{1}{(1-r_{max})^2} = 4, \tag{3.81}$$

that is, no more than 3 bits can be optimally encoded in the qubit. This result has been derived before by both geometric [87] and non-locality [88] arguments, but our derivation relies completely on simplex embedding. We believe that deriving bounds on quantum communication tasks purely from simplex embedding arguments is a very powerful consequence of adopting contextuality as the main notion of nonclassicality.

3.5.4 Robustness to dephasing in the 3-to-1 POM task

Finally, we consider a family of 3-to-1 POM scenarios parametrised by an angle θ, in the spirit of Ref. [Third Paper]. The preparations are given by

$$\begin{aligned}\Omega^F \quad := \quad & \left\{ \frac{1}{2}\begin{pmatrix} 1+\cos\theta & e^{\pm i\frac{\pi}{4}}\sin\theta \\ e^{\mp i\frac{\pi}{4}}\sin\theta & 1-\cos\theta \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1-\cos\theta & -e^{\pm i\frac{\pi}{4}}\sin\theta \\ -e^{\mp i\frac{\pi}{4}}\sin\theta & 1+\cos\theta \end{pmatrix}, \right. \\ & \left. \frac{1}{2}\begin{pmatrix} 1+\cos\theta & -e^{\pm i\frac{\pi}{4}}\sin\theta \\ -e^{\mp i\frac{\pi}{4}}\sin\theta & 1-\cos\theta \end{pmatrix}, \frac{1}{2}\begin{pmatrix} 1-\cos\theta & e^{\pm i\frac{\pi}{4}}\sin\theta \\ e^{\mp i\frac{\pi}{4}}\sin\theta & 1+\cos\theta \end{pmatrix} \right\}, \end{aligned} \tag{3.82}$$

and the effects are given by the eigenvectors of the three Pauli operators X, Y and Z.

First, we input these states and effects to the linear program from Ref. [Second Paper] to obtain the plot in the left in Figure 3.1. Notice that, as expected, robustness of contextuality to depolarising grows proportionally with the success rate. Furthermore, this plot is recovered by Eq. 3.80 when $s_{NC} = \frac{2}{3}$, the noncontextual success rate for the 3-to-1 POM task.

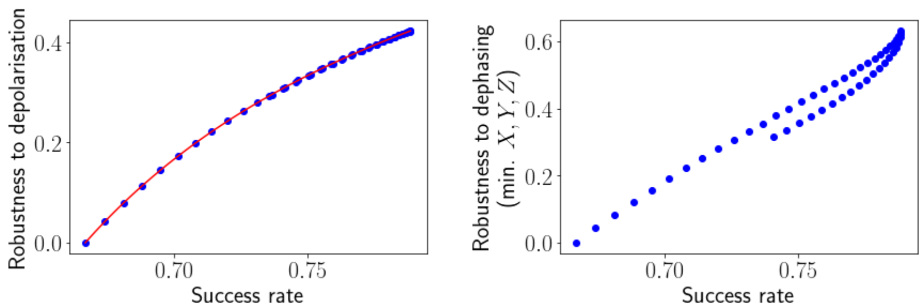


Figure 3.1: (Left) Robustness of contextuality to depolarising noise vs. success rate for the 3-to-1 POM task. Red curve reproduces Eq. 3.80. (Right) Minimal robustness of contextuality to dephasing noise across the measurements available in the fragment vs. success rate for the 3-to-1 POM task.

Moreover, we employ the modified code to explore the action of dephasing on the same scenario, obtaining a plot that does not display the same monotonic proportionality. We show however that by minimizing the robustness of contextuality to dephasing over all relevant dephasing basis a proportionality is again observed between the monotonicity of this quantity and of the success rate for the task. In particular, this minimisation is implemented across the X, Y and Z axes of the Bloch sphere, and we provide arguments on why only these axes are the relevant ones in Ref. [Preprint]. This plot is given in the right side of Figure 3.1. Evidently, these quantifiers are not as consistent as for the case of depolarisation. However, we argue that this minimised robustness to dephasing is still a good quantifier since (i) it captures the same extremal cases as success rate and (ii) knowing the value of the minimised robustness bounds the possible success rates one can obtain in the task.

We believe robustness to be a good operational measure of nonclassicality, and our work also shows that it can be a good quantifier for practical tasks in which contextuality is the source of nonclassical advantage. Moreover, all the conclusions in this work are drawn partially from numerical calculations obtained with the linear program from Ref. [Second Paper] and Ref. [Repository], displaying the practicality of this tool for both foundational and information-theoretical purposes.

Chapter 4

Outlook

This thesis explores how assuming contextuality as the main notion of nonclassicality for physical theories can help on the characterisation of nonclassical (and even beyond-quantum) behaviours in different prepare-and-measure scenarios. Contextuality has been demonstrated to subsume or be related to many other signatures of nonclassicality, and I reproduced in Chapter 2 some of the compelling arguments that justify its physical appeal. In particular, I approach contextuality from two frameworks: the hypergraph approach and the GPT framework approach. With the first approach, I visit EPR scenarios to characterise sets of assemblages from contextuality principles, and show that we need stronger device-dependent principles for such. I then move to the GPT framework, which is naturally a device-dependent description of an operational scenario, and introduce a linear program for assessing contextuality in prepare-and-measure scenarios. This linear program and its implementation are the highlight of this thesis, and I showcase its usefulness in prepare-and-measure scenarios related to minimum-error discrimination and parity-oblivious multiplexing tasks. In the first, I report how we employ the code to investigate the relation between contextuality and coherence, concluding that there are scenarios that display contextuality with any amount of coherence and in a maximally robust way. In the second, I comment on how robustness of contextuality can work as a quantifier of quantum advantage, and how the linear program can play a role on recovering bounds for optimal encoding.

These results showcase the potential of the GPT framework to study contextuality by exploring the numerical and analytical tools provided by this framework to characterise quantum advantage in different tasks. Concerning macroscopic noncontextuality, the notion of macrorealism has recently been formulated in the GPT framework [17]. Can we formalise macroscopic realisations of GPTs in a similar spirit to Ref. [First Paper]? If so, does imposing macrorealism or simplex-embeddability over the macroscopic GPTs imply any significant constraint on the original GPTs? Moreover, some attention in the literature has been devoted to situations in which assessments of contextuality can fail [36, 42, 61, 62, 86]. An ongoing project of our group has been exploring under which circumstances assessments of contextuality are trustworthy even if one has not access to tomographic sets of states and effects. Morevoer, the linear program can miss any contextuality present in transformations, something that was originally investigated in Spekkens' seminal work [11]. Could we extend the characterisation of positive-cones to transformations, assessing their simplex-embeddability? Concerning the modification of the program to assess robustness to dephasing, our most recent formulation is sufficient only for strongly self-dual GPTs. It would be interesting to define a notion of dephasing noise for GPTs beyond those, or at least investigate which properties such noise is expected to have in these beyond-quantum contexts.

Bibliography

1. Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical review*, 47(10):777, 1935. (cited on pages 1 and 15)

2. E. Schrödinger. Probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 32(3):446–452, 1936. (cited on pages 1 and 15)

3. J. S. Bell. On the einstein podolsky rosen paradox. *Phys. Phys. Fiz.*, 1:195–200, Nov 1964. (cited on page 1)

4. J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed Experiment to Test Local Hidden-Variable Theories. *Phys. Rev. Lett.*, 23(15):880–884, Oct 1969. (cited on page 1)

5. S. Kochen and E. P. Specker. The Problem of Hidden Variables in Quantum Mechanics. *Journal Mathematics and Mechanics*, (vol 17 N. 1), 1967. (cited on pages 1 and 7)

6. Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Rev. Mod. Phys.*, 86:419–478, Apr 2014. (cited on pages 1 and 15)

7. Costantino Budroni, Adán Cabello, Otfried Gühne, Matthias Kleinmann, and Jan-Åke Larsson. Kochen-specker contextuality. *Rev. Mod. Phys.*, 94:045007, Dec 2022. (cited on page 1)

8. Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. *Rev. Mod. Phys.*, 81:865–942, Jun 2009. (cited on pages 1 and 15)

9. A. Streltsov, G. Adesso, and M. B. Plenio. "*colloquium*: quantum coherence as a resource". *Rev. Mod. Phys.*, 89:041003, 2017. (cited on page 1)

10. Anindita Bera, Tamoghna Das, Debasis Sadhukhan, Sudipto Singha Roy, Aditi Sen(De), and Ujjwal Sen. Quantum discord and its allies: a review of recent progress. *Reports on Progress in Physics*, 81(2):024001, December 2017. (cited on page 1)

11. R. W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Phys. Rev. A*, 71:052108, May 2005. (cited on pages 1, 7, and 33)

12. R. W. Spekkens. "The ontological identity of empirical indiscernibles: Leibniz's methodological principle and its significance in the work of Einstein". *arXiv:1909.04628 [physics.hist-ph]*, 2019. (cited on pages 1 and 7)

13. Y. C. Liang, R. W. Spekkens, and H. M. Wiseman. “Specker’s parable of the over-protective seer: A road to contextuality, nonlocality and complementarity”. *Physics Reports*, 506:1–39, 2011. (cited on pages 1 and 7)

14. V. Wright and M. Farkas. “An invertible map between Bell non-local and contextuality scenarios”. *arXiv:2211.12550 [quant-ph]*, 2022. (cited on page 7)

15. R. D. Baldijão, R. Wagner, C. Duarte, B. Amaral, and M. T. Cunha. “Emergence of Noncontextuality under Quantum Darwinism”. *PRX Quantum*, 2:030351, 2021. (cited on page 7)

16. R. Kunjwal, M. Lostaglio, and M. F. Pusey. “Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts”. *Phys. Rev. A*, 100:042116, 2019. (cited on page 7)

17. D. Schmid. “Macrorealism as strict classicality in the framework of generalized probabilistic theories (and how to falsify it)”. *arXiv:2209.11783 [quant-ph]*, 2022. (cited on pages 7 and 33)

18. R. W. Spekkens. “Negativity and Contextuality are Equivalent Notions of Nonclassicality”. *Phys. Rev. Lett.*, 101:020401, 2008. (cited on page 7)

19. D. Schmid, J. H. Selby, and R. W. Spekkens. “Unscrambling the omelette of causation and inference: The framework of causal-inferential theories”. *arXiv:2009.03297 [quant-ph]*, 2020. (cited on page 12)

20. David Schmid, Roberto D. Baldijão, Yìlè Yīng, Rafael Wagner, and John H. Selby. Kirkwood-dirac representations beyond quantum states (and their relation to noncontextuality). *arXiv:2405.04573 [quant-ph]*, 2024. (cited on pages 1 and 7)

21. D. Saha and A. Chaturvedi. “Preparation contextuality as an essential feature underlying quantum communication advantage”. *Phys. Rev. A*, 100:022108, 2019. (cited on pages 1 and 7)

22. Abhyoudai S. S., Sumit Mukherjee, and A. K. Pan. Robust certification of unsharp instruments through sequential quantum advantages in a prepare-measure communication game. *Phys. Rev. A*, 107:012411, Jan 2023.

23. Carles Roch i Carceller, Kieran Flatt, Hanwool Lee, Joonwoo Bae, and Jonatan Bohr Brask. Quantum vs noncontextual semi-device-independent randomness certification. *Phys. Rev. Lett.*, 129:050501, Jul 2022. (cited on pages 1 and 7)

24. D. Schmid, H. Du, J. H. Selby, and M. F. Pusey. “Uniqueness of Noncontextual Models for Stabilizer Subtheories”. *Phys. Rev. Lett.*, 129, 2022. (cited on pages 1 and 7)

25. Robert W. Spekkens, D. H. Buzacott, A. J. Keehn, Ben Toner, and G. J. Pryde. Preparation contextuality powers parity-oblivious multiplexing. *Phys. Rev. Lett.*, 102:010401, Jan 2009. (cited on pages 1, 7, 30, and 31)

26. A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai. “Parity Oblivious d-Level Random Access Codes and Class of Noncontextuality Inequalities”. *arXiv:1607.05490 [quant-ph]*, 2016.

27. A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora. "Optimal bounds for parity-oblivious random access codes". *New J. Phys.*, 18:045003, 2016. (cited on pages 30 and 31)

28. S. A. Yadavalli and R. Kunjwal. "Contextuality in entanglement-assisted one-shot classical communication". *Quantum*, 6:839, 2022. (cited on pages 1 and 7)

29. Joseph Bowles, Victoria J Wright, Máté Farkas, Nathan Killoran, and Maria Schuld. Contextuality and inductive bias in quantum machine learning, 2023. (cited on pages 1 and 7)

30. M. Lostaglio. "Certifying Quantum Signatures in Thermodynamics and Metrology via Contextuality of Quantum Linear Response". *Phys. Rev. Lett.*, 125:230603, 2020.

31. M. Lostaglio and G. Senno. "Contextual advantage for state-dependent cloning". *Quantum*, 4:258, 2020.

32. D. Schmid and R. W. Spekkens. "Contextual Advantage for State Discrimination". *Phys. Rev. X*, 8:011015, 2018. (cited on pages 2, 7, 11, 24, and 25)

33. J. Shin, D. Ha, and Y. Kwon. "Quantum Contextual Advantage Depending on Nonzero Prior Probabilities in State Discrimination of Mixed Qubit States". *Entropy*, 23:1583, 2021.

34. K. Flatt, H. Lee, C. Roch I Carceller, J. B. Brask, and J. Bae. "Contextual Advantages and Certification for Maximum-Confidence Discrimination". *PRX Quantum*, 3:030337, 2022.

35. S. Mukherjee, S. Naonit, and A. K. Pan. "Discriminating three mirror-symmetric states with a restricted contextual advantage". *Phys. Rev. A*, 106:012216, 2022. (cited on pages 1 and 7)

36. David Schmid, John H. Selby, and Robert W. Spekkens. Addressing some common objections to generalized noncontextuality. *Phys. Rev. A*, 109:022228, Feb 2024. (cited on pages 1, 11, 18, and 33)

37. Miguel Navascués, Stefano Pironio, and Antonio Acín. Bounding the set of quantum correlations. *Physical Review Letters*, 98(1):010401, 2007. (cited on page 2)

38. Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008.

39. Miguel Navascués, Yelena Guryanova, Matty J Hoban, and Antonio Acín. Almost quantum correlations. *Nature communications*, 6(1):1–7, 2015. (cited on pages 2, 9, and 17)

40. David Schmid, Robert W. Spekkens, and Elie Wolfe. All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences. *Phys. Rev. A*, 97:062103, Jun 2018. (cited on page 2)

41. Anubhav Chaturvedi, Máté Farkas, and Victoria J Wright. Characterising and bounding the set of quantum behaviours in contextuality scenarios. *Quantum*, 5:484, June 2021.

42. Victor Gitton and Mischa P Woods. On the system loophole of generalized noncontextuality. 2022. (cited on pages 2, 11, and 33)

43. Howard M Wiseman, Steve James Jones, and Andrew C Doherty. Steering, entanglement, nonlocality, and the Einstein-Podolsky-Rosen paradox. *Physical review letters*, 98(14):140402, 2007. (cited on pages 2 and 15)

44. R. W. Spekkens. “Evidence for the epistemic view of quantum states: A toy theory”. *Phys. Rev. A*, 75:032110, 2007. (cited on pages 2, 22, 24, and 27)

45. L. Hardy. “Quantum Theory From Five Reasonable Axioms”. *arXiv:quant-ph/0101012*, 2001. (cited on pages 5 and 9)

46. J. Barrett. “Information processing in generalized probabilistic theories”. *Phys. Rev. A*, 75:032304, 2007.

47. Giacomo Mauro D’Ariano, Giulio Chiribella, and Paolo Perinotti. *Quantum Theory from First Principles: An Informational Approach*. Cambridge University Press, 2017. (cited on page 5)

48. David Schmid, John H. Selby, Elie Wolfe, Ravi Kunjwal, and Robert W. Spekkens. Characterization of noncontextuality in the framework of generalized probabilistic theories. *PRX Quantum*, 2:010331, Feb 2021. (cited on pages 7, 11, and 12)

49. Antonio Acín, Tobias Fritz, Anthony Leverrier, and Ana Belén Sainz. A combinatorial approach to nonlocality and contextuality. *Communications in Mathematical Physics*, 334(2):533–628, January 2015. (cited on pages 8, 9, and 18)

50. Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, March 1994. (cited on page 8)

51. M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Żukowski. Information causality as a physical principle. *Nature*, 461(7267):1101–1104, October 2009. (cited on pages 8 and 17)

52. Miguel Navascués and Harald Wunderlich. A glance beyond the quantum model. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 466(2115):881–890, November 2009. (cited on page 17)

53. T. Fritz, A.B. Sainz, R. Augusiak, J Bohr Brask, R. Chaves, A. Leverrier, and A. Acín. Local orthogonality as a multipartite principle for quantum correlations. *Nature Communications*, 4:2263, 2013.

54. A. B. Sainz, T. Fritz, R. Augusiak, J. Bohr Brask, R. Chaves, A. Leverrier, and A. Acín. Exploring the local orthogonality principle. *Phys. Rev. A*, 89:032117, Mar 2014.

55. Joe Henson and Ana Belén Sainz. Macroscopic noncontextuality as a principle for almost-quantum correlations. *Phys. Rev. A*, 91:042114, Apr 2015. (cited on page 18)

56. Patricia Contreras-Tejada, Giannicola Scarpa, Aleksander M Kubicki, Adam Brandenburger, and Pierfrancesco La Mura. Observers of quantum systems cannot agree to disagree. *Nature communications*, 12(1):1–7, 2021. (cited on pages 8 and 17)

57. G. Chiribella, G. M. D’Ariano, and P. Perinotti. “Probabilistic theories with purification”. *Phys. Rev. A*, 81:062348, 2010. (cited on page 9)

58. John H. Selby, Carlo Maria Scandolo, and Bob Coecke. Reconstructing quantum theory from diagrammatic postulates. *Quantum*, 5:445, April 2021. (cited on page 9)

59. Bob Coecke and Aleks Kissinger. *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning.* Cambridge University Press, March 2017. (cited on page 10)

60. John H. Selby, David Schmid, Elie Wolfe, Ana Belén Sainz, Ravi Kunjwal, and Robert W. Spekkens. Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing nonclassicality. *Phys. Rev. A*, 107:062203, Jun 2023. (cited on pages 10 and 19)

61. Matthew F. Pusey, Lídia del Rio, and Bettina Meyer. Contextuality without access to a tomographically complete set, 2019. (cited on pages 11 and 33)

62. Victor Gitton and Mischa P. Woods. Solvable criterion for the contextuality of any prepare-and-measure scenario. *Quantum*, 6:732, June 2022. (cited on pages 11 and 33)

63. R. Kunjwal and R. W. Spekkens. "From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism". *Phys. Rev. Lett.*, 115:110403, 2015. (cited on page 11)

64. I. Marvian. "Inaccessible information in probabilistic models of quantum systems, non-contextuality inequalities and noise thresholds for contextuality". *arXiv:2003.05984 [quant-ph]*, 2020. (cited on page 11)

65. Ana Belén Sainz, Nicolas Brunner, Daniel Cavalcanti, Paul Skrzypczyk, and Tamás Vértesi. Postquantum steering. *Physical review letters*, 115(19):190403, 2015. (cited on pages 16 and 18)

66. Nicolas Gisin. Stochastic quantum dynamics and relativity. *Helvetica Physica Acta*, 62(4):363–371, 1989. (cited on page 16)

67. Lane P Hughston, Richard Jozsa, and William K Wootters. A complete classification of quantum ensembles having a given density matrix. *Physics Letters A*, 183(1):14–18, 1993. (cited on page 16)

68. Michael J. Grabowecky, Christopher A. J. Pollack, Andrew R. Cameron, Robert W. Spekkens, and Kevin J. Resch. Experimentally bounding deviations from quantum theory for a photonic three-level system using theory-agnostic tomography. *Phys. Rev. A*, 105:032204, Mar 2022. (cited on page 18)

69. Elie Wolfe. `SimplexEmbedding`: Tools to discover a gpt's simplicial (cone) embedding. https://github.com/eliewolfe/SimplexEmbedding, 2022. (cited on pages 19 and 29)

70. Daniel Gottesman. Stabilizer codes and quantum error correction, 1997. (cited on page 19)

71. Anthony J Short and Jonathan Barrett. Strong nonlocality: a trade-off between states and measurements. *New Journal of Physics*, 12(3):033034, March 2010. (cited on page 19)

72. Peter Janotta and Raymond Lal. Generalized probabilistic theories without the no-restriction hypothesis. *Phys. Rev. A*, 87:052131, May 2013. (cited on pages 19 and 29)

73. Martin Plávala. General probabilistic theories: An introduction. *Phys. Rep.*, 1033:1–64, September 2023. (cited on page 19)

74. J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens. "Contextuality without incompatibility". *arXiv:2106.09045 [quant-ph]*, 2021. (cited on page 19)

75. Günter M. Ziegler. *Shellability and the Upper Bound Theorem*, pages 231–290. Springer New York, New York, NY, 1995. (cited on page 19)

76. Theodore S Motzkin, Howard Raiffa, Gerald L Thompson, and Robert M Thrall. The double description method. *Contributions to the Theory of Games*, 2(28):51–73, 1953. (cited on page 20)

77. Jingyan Lin, Kunkun Wang, Lei Xiao, Dengke Qu, Gaoyan Zhu, Yong Zhang, and Peng Xue. Experimental investigation of contextual robustness and coherence in state discrimination. *Phys. Rev. A*, 109:052208, May 2024. (cited on page 23)

78. Taira Giordani, Rafael Wagner, Chiara Esposito, Anita Camillini, Francesco Hoch, Gonzalo Carvacho, Ciro Pentangelo, Francesco Ceccarelli, Simone Piacentini, Andrea Crespi, Nicolò Spagnolo, Roberto Osellame, Ernesto F. Galvão, and Fabio Sciarrino. Experimental certification of contextuality, coherence, and dimension in a programmable universal photonic processor. *Science Advances*, 9(44):eadj4249, 2023. (cited on page 23)

79. C. W. Helstrom. "Quantum detection and estimation theory". *J. Stat. Phys.*, 1:231, 1969. (cited on page 25)

80. Mathew Weiss. `pySimplexEmbedding`: Tools to discover a gpt's simplicial (cone) embedding (in python!). https://github.com/heyredhat/pySimplexEmbedding, 2022. (cited on page 28)

81. Peter Janotta, Christian Gogolin, Jonathan Barrett, and Nicolas Brunner. Limits on nonlocal correlations from the structure of the local state space. *New Journal of Physics*, 13(6):063024, June 2011. (cited on page 29)

82. Markus P. Müller and Cozmin Ududec. Structure of reversible computation determines the self-duality of quantum theory. *Phys. Rev. Lett.*, 108:130401, Mar 2012. (cited on page 29)

83. A. S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems Inform. Transmission*, 9:177–183, 1973. (cited on page 30)

84. Gilles Brassard. *Foundations of Physics*, 33(11):1593–1616, 2003. (cited on page 30)

85. Nicolas Gisin and Rob Thew. Quantum communication. *Nature Photonics*, 1(3):165–171, March 2007. (cited on page 30)

86. Lorenzo Catani, Thomas D. Galley, and Tomáš Gonda. Resource-theoretic hierarchy of contextuality for general probabilistic theories, 2024. (cited on pages 30 and 33)

87. M Hayashi, K Iwama, H Nishimura, R Raymond, and S Yamashita. (4,1)-quantum random access coding does not exist—one qubit is not enough to recover one of four bits. *New Journal of Physics*, 8(8):129, aug 2006. (cited on page 31)

88. Shouvik Ghorai and A. K. Pan. Optimal quantum preparation contextuality in an n-bit parity-oblivious multiplexing task. *Phys. Rev. A*, 98:032110, Sep 2018. (cited on page 31)

Journal of Physics A: Mathematical and Theoretical

PAPER

On characterising assemblages in Einstein–Podolsky–Rosen scenarios

IOP Publishing Journal of Physics A: Mathematical and Theoretical

J. Phys. A: Math. Theor. **55** (2022) 264002 (22pp) https://doi.org/10.1088/1751-8121/ac7090

On characterising assemblages in Einstein–Podolsky–Rosen scenarios

Vinicius P Rossi[1,*], **Matty J Hoban**[2,3] **and Ana Belén Sainz**[1]

[1] International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-308 Gdańsk, Poland

[2] Quantinuum LLC and Cambridge Quantum Computing Ltd, Cambridge, United Kingdom

[3] Department of Computing, Goldsmiths, University of London, New Cross, London SE14 6NW, United Kingdom

E-mail: vinicius.prettirossi@phdstud.ug.edu.pl

Received 14 February 2022
Accepted for publication 17 May 2022
Published 6 June 2022

Abstract

Characterising non-classical quantum phenomena is crucial not only from a fundamental perspective, but also to better understand its capabilities for information processing and communication tasks. In this work, we focus on exploring the characterisation of Einstein–Podolsky–Rosen inference (a.k.a. steering): a signature of non-classicality manifested when one or more parties in a Bell scenario have their systems and measurements described by quantum theory, rather than being treated as black boxes. We propose a way of characterising common-cause assemblages from the correlations that arise when the trusted party performs tomographically-complete measurements on their share of the experiment, and discuss the advantages and challenges of this approach. Within this framework, we show that so-called almost quantum assemblages satisfy the principle of macroscopic noncontextuality, and demonstrate that a subset of almost quantum correlations recover almost quantum assemblages in this approach.

Keywords: Einstein–Podolsky–Rosen steering scenarios, foundations of quantum theory, characterising quantum theory from principles, almost-quantum correlations

Contents

1. Introduction 2
2. Einstein–Podolsky–Rosen scenarios: preliminaries 4

*Author to whom any correspondence should be addressed.

2.1. Bipartite EPR scenarios 4
2.2. Assemblages without a quantum realisation 5
2.3. Multipartite EPR scenarios 6
2.4. Almost-quantum assemblages and moment matrices 7
3. Characterising assemblages 10
3.1. Bipartite EPR scenarios 11
4. Characterising assemblages by characterising correlations 13
4.1. Macroscopic classicality considerations 14
4.2. The need for device-dependent principles 15
4.3. A stronger device-dependent version of macroscopic non-contextuality 16
5. Discussion 18
Acknowledgments 19
Data availability statement 19
References 20

1. Introduction

Identifying the non-classical aspects of quantum theory is not only an important research theme in the foundations of quantum theory, but it is now of relevance to quantum computation and quantum information. In particular, these non-classical aspects can be seen as a resource that can be used for a quantum advantage in computation [1–4] or secure communication [5–7]. Of special interest is the study of non-classical resources in Bell scenarios, often termed 'Bell non-locality', which is a resource in protocols for random number generation [8–10], quantum key distribution [6, 7], and verifiable, delegated quantum computation [11]. Interestingly, the protocols where this non-classicality is a resource are *device-independent*, which means that (possibly) quantum devices can be treated as multi-party *black boxes* that have classical inputs and outputs: one does not need to specify *a priori* the quantum description of the devices inside the black box. In the nomenclature of reference [12], these resources can be termed *common-cause boxes*, which comes from the structure of a Bell experiment [13].

Importantly, when considering resources, one must understand their limitations as well as their possibilities for enhancing technological performance. The characterisation of the set of common-cause boxes possible in quantum theory is useful for understanding the scope of device-independent information processing since it can, for instance, limit the ability of an eavesdropper to predict the output of a black box and thus lead to private randomness generation [8]. One important example where we see the limitations of quantum systems is in the Tsirelson bound [14], which dictates the largest violation possible with quantum common-cause boxes of the Clauser–Horne–Shimony–Holt inequality [15]. The impact of this limitation includes that Alice and Bob cannot use quantum common-cause boxes to enhance communication so to transmit arbitrary messages with only one bit of classical communication [16]. In turn, quantum common-cause boxes that give this maximal violation can be used for all of the kinds of the aforementioned device-independent protocols.

Curiously, the Tsirelson bound is not the largest numerical violation possible, and such *post-quantum* violations can be achieved with common-cause boxes that respect the no-signalling principle: inputs for one party do not influence the statistics of another party's output. Popescu and Rohrlich used this example (now called the PR box) to point out that for these boxes, special relativity is not enough to constrain one to the set of quantum boxes, and asked what axioms or principles could single out this set [17]. Answering such a question would give us deep insight into why quantum theory is one of our ultimate theories, and explain its limitations for information processing.

Subsequent to the work of Popescu and Rohrlich, van Dam described an *information theoretic* principle that ruled out the PR box: the principle of non-trivial communication complexity [16]. Despite follow-up work by Brassard *et al* [18] it was unclear if this principle could recover quantum common-cause boxes. Significant progress on this question was made by the seminal work of Navascués, Pironio and Acín (NPA) [19–21], whom developed a hierarchy of semi-definite programs that converge to a set of quantum common-cause boxes[5].

The NPA hierarchy is an extremely useful tool, but is not described in terms of either information-theoretic or physical principles. As a result, many principles followed such as information causality [22], macroscopic locality [23], local orthogonality [24, 25], and no common certainty of disagreement [26]. An obstacle to the development of such principles is the almost-quantum set of common-cause boxes [27], which have not yet violated any of the proposed principles.

Independently of the study of common-cause boxes, another line of research in this topic pertains to the characterisation of not only common-cause boxes, but of the whole of quantum theory. Unlike for common-case boxes, various axiomatisations of (finite-dimensional) quantum theory have been successfully developed [28–45], especially in the study of *generalised probabilistic theories* [46] or *process theories* [47]. Furthermore, tentative progress has been made in understanding the information processing power of quantum theory within such a broad framework [46, 48]. The starting point for these frameworks requires one, in some sense, to dictate the degrees of freedom in a general, possibly non-quantum, experiment from the outset so it can be associated with some vector space. This approach is thus very different from the device-independent approach as it requires a characterisation of the systems involved. As a result, this approach has not led to deep intuitions about the strengths or limitations of quantum phenomena such as non-classical common-cause boxes.

These two approaches to characterising quantum systems within a more general framework—a.k.a. characterising quantum theory 'from the outside'—are not the only possibilities. In particular, one situation is to start with some elements of quantum theory, and then go beyond this to allow something more general. A notable example in this direction is the development of *process matrices* that exhibit indefinite causal order [49]. Other examples include the study of Bell scenarios assuming only local measurements [50–52] and recovering the measurement postulates of quantum theory from just unitary dynamics [53]. In this work, we also pursue a direction of assuming some aspect of quantum theory, in particular we assume that quantum theory holds locally for a single party alone.

To be more precise, instead of considering Bell scenarios, we consider Einstein–Podolsky–Rosen (EPR) scenarios [54, 55], inspired by their seminal 1935 paper [56]. In such a scenario, as mentioned, there is a single party (called Bob) that has a quantum system with known degrees of freedom (with a particular Hilbert space dimension) that can be directly measured; thus the system has a quantum-theoretic description. Outside of Bob's laboratory we can treat everything else as uncharacterised, but satisfying causal constraints. In particular, there are multiple, non-communicating parties that generate classical data. In this way we can see the EPR scenario as a Bell scenario with these extra assumptions

[5] Technically, the hierarchy converges to the set of common-cause boxes realised by commuting measurements on a quantum state, which, for infinite dimensional systems, is distinct from boxes produced by local measurements defined according to a tensor product. For finite dimensional quantum systems, the two definitions are equivalent.

made about one particular party. While in Bell scenarios one considers boxes associated with certain correlations, in EPR scenarios the fundamental object is an *assemblage* [57]. An assemblage is a collection of (sub-normalised) conditional density matrices: a description of the quantum system in Bob's laboratory conditioned on the classical data generated outside of this laboratory. An assemblage provides a rich structure since it can capture both box-like correlations as well as quantum information.

Non-classical[6] assemblages in EPR scenarios[7] show a particular non-classical aspect of nature that quantum theory features in the lab. In addition, they have been shown useful for enhancing our performance at cryptographic protocols [5, 58–60], hence manifesting as useful quantum resources. To capture this non-classical resource one needs to characterise the set of classical assemblages, and much work has been devoted to this [60–62]. However, the question of 'how to characterise non-classical assemblages from the outside' has barely been explored. That is, given a general framework for non-classical assemblages can we single out those with a completely quantum explanation from those that go beyond the quantum formalism? As well as being fundamental, this question gives us another avenue to explore the resourcefulness of non-classical assemblages. In this work we comment on the challenges we face when tackling this question, on a particular way to overcome these, and on the new insights from and limitations of our approach.

2. Einstein–Podolsky–Rosen scenarios: preliminaries

EPR scenarios correspond to a particular type of experiment where non-classical features of nature manifest, and captures a particular type of non-classical property referred to as *EPR inference*[8]. Some features in the specification of an EPR scenario are similar to those which specify Bell scenarios, so the reader familiar with the latter might find a similar discourse below. It is worth mentioning that in this paper we focus on EPR inference of finite-dimensional quantum systems, that is, Bob's system (see below) has a finite dimension.

For simplicity, let us first introduce the case of two parties, hereon called Alice and Bob (see figure 1). Alice and Bob share a physical system prepared on a state possibly unknown to them, and Alice wants to learn about the system held by Bob in his lab. Alice and Bob are distant from each other and cannot communicate (neither classically nor with quantum systems), hence the only way that Alice can learn about the state of Bob's system is by performing measurements on her share of the system and making inferences about Bob's. After performing a measurement, then, Alice can update her knowledge on the state preparation of Bob's quantum system conditioned on her obtained outcome. In this sense, each measurement choice by Alice selects an ensemble of the possible updated states of Bob's quantum system, and the probability of each of her outcomes yields the probability with which each of those updated states will arise. The collection of these ensembles is called *assemblage* and is the object of interest in the study of EPR scenarios. Notice that this discourse is different from the ubiquitous one where EPR scenarios are about non-classical causal influences, the famous 'spooky action at a distance'. For a deeper discussion on the advantages of taking this less-common inferential perspective we refer the reader to reference [63].

[6] In the literature, these assemblages are also referred to as 'steerable' assemblages.

[7] In the literature, these are also known as 'steering' scenarios.

[8] In the literature, EPR inference is also known as 'EPR steering' or merely 'steering'.

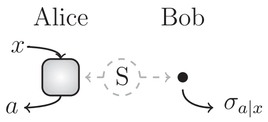


Figure 1. Bipartite EPR scenario: Alice refines her knowledge about the state of Bob's quantum system by making inferences from performing measurements on her share of the system. Alice's measurement choices are labelled by the classical variable x that takes values in the set $\mathbb{X}$, and the measurement outcomes are labelled by the classical variable a that takes values in the set $\mathbb{A}$. The possibly subnormalised positive-semidefinite matrix $\sigma_{a|x}$ represents the state of Bob's system when Alice's measurement choice and outcome are x and a respectively. The probability that Alice obtains a when performing measurement x is given by $p(a|x) = \text{tr}\left\{\sigma_{a|x}\right\}$. The assemblage (ensemble of ensembles) representing the information on the quantum state of Bob's system is $\mathbf{\Sigma}_{\mathbb{A}|\mathbb{X}} = \{\{\sigma_{a|x}\}_{a\in\mathbb{A}}\}_{x\in\mathbb{X}}$.

2.1. Bipartite EPR scenarios

In a bipartite scenario, we denote by $\mathbb{X}$ the set of classical labels for Alice's measurement choices, and $\mathbb{A}$ the set of classical labels for her measurement outcomes[9]. In addition, let us denote by $\mathcal{H}_B$ the Hilbert space corresponding to Bob's quantum system, which is moreover known to both parties. An assemblage, denoted by $\mathbf{\Sigma}_{\mathbb{A}|\mathbb{X}}$, is hence given by the collection of (possibly unnormalised) quantum states $\mathbf{\Sigma}_{\mathbb{A}|\mathbb{X}} = \{\{\sigma_{a|x}\}_{a\in\mathbb{A}}\}_{x\in\mathbb{X}}$, where $p(a|x) = \text{tr}\left\{\sigma_{a|x}\right\}$ is the probability that Alice obtains outcome $a \in \mathbb{A}$ when performing measurement $x \in \mathbb{X}$, and $\rho_{a|x} = \frac{\sigma_{a|x}}{p(a|x)}$ is the normalised quantum state that refines Alice's knowledge on the state of Bob's quantum system.

An assemblage $\mathbf{\Sigma}_{\mathbb{A}|\mathbb{X}}$ is said to admit a *quantum realisation* if there exists a Hilbert space $\mathcal{H}_A$ for Alice, a set of measurements $\{\{M_{a|x}\}_{a\in\mathbb{A}}\}_{x\in\mathbb{X}}$ in $\mathcal{H}_A$, and a quantum state ρ in $\mathcal{H}_A \otimes \mathcal{H}_B$, such that the elements of $\mathbf{\Sigma}_{\mathbb{A}|\mathbb{X}}$ can be expressed as:

$$\sigma_{a|x} = \text{tr}_A\left\{M_{a|x} \otimes \mathbb{I}_{\mathcal{H}_B}\rho\right\} \quad \forall\, a \in \mathbb{A},\ x \in \mathbb{X}. \tag{1}$$

2.2. Assemblages without a quantum realisation

In this work we are interested in characterising the set of quantum-realisable assemblages by singling them out within a broader class of assemblages that may include non-quantum-realisable ones (known as post-quantum assemblages). Two comments are in order for understanding this question: how may a post-quantum assemblage arise within such a seemingly-quantum-exclusive experiment? And how are post-quantum assemblages mathematically specified?

Regarding the first question, there are different ways in which one can capture what is necessarily quantum in this experiment, and what aspects of it may be dictated by a beyond-quantum theory. Informally, one may think of a Universe that can be fully fleshed out locally using a quantum description, but that holistically requires something beyond the quantum formalism to be described and understood. Alternatively, one can also think of the constraints of an EPR scenario as just one more way to probe nature, and explore the scope of lessons to learn from

[9] In principle, different measurement can have different number of outcomes, but for the current discussion we can take these sets to be the same (and all equal to $\mathbb{A}$) without loss of generality.

it. This latter approach takes on the device-independent approach underpinning a Bell scenario, where even if the parties had access to a quantum system and measurements, the only information that they rely on to assess the properties of nature are the classical labels of the measurement choices and outcomes, and the statistics one may draw from them (no need to assume or use the quantum formalism at all!). So, given that an EPR scenario is defined by $\mathbb{X}$, $\mathbb{A}$, and $\mathcal{H}_B$, then considerations beyond quantum theory may come in in the nature of the system shared by Alice and Bob. This perspective goes along the lines of a causal grounding of an EPR scenario, and gauges the classicality or quantumness of an assemblage in terms of the common causes (which may be thought of as the shared system) that correlate Alice and Bob's labs. Another possibility comes from the constraint that in an EPR scenario Alice and Bob are distant parties that perform space-like separated local actions and cannot communicate with each other. This also opens the door for some non-quantumness to sneak in the form of fine-tuned hidden signalling. However, given the conceptual problems that fine-tuning comes with [64] we encourage the reader not to entertain this latter approach. Nevertheless, a perspective provided in terms of imposing an overall no-signalling principle between Alice and Bob allows one to include in the picture assemblages with no quantum realisation [65–68], in the same way that in Bell scenarios post-quantum correlations (such as the one given by Popescu and Rohrlich [17]) are encompassed.

Regardless of which fundamental way one interprets the possibility of there being EPR inferences that have no quantum explanation, there is the formal question of how are general assemblages (be it quantum or post-quantum) mathematically specified. Luckily, the two ways described above (common-cause perspective and no-signalling perspective) define the same set of mathematical objects as their most general set of assemblages [69]. Hence, one can readily explore the consequences of post-quantum EPR inferences while mulling over its philosophical implications. In the case of a bipartite EPR scenario, the definition of a general assemblage is as follows:

Definition 1. Common-cause assemblage, a.k.a. non-signalling assemblage.

Consider a bipartite EPR scenario, where Alice's measurements and outcomes are labelled by the elements of the sets $\mathbb{X}$ and $\mathbb{A}$, respectively, and Bob's quantum system is represented by the Hilbert space $\mathcal{H}_B$. Then, an assemblage $\mathbf{\Sigma}_{\mathbb{A}|\mathbb{X}}$ is a common-cause assemblage (equivalently, a non-signalling assemblage) if the following constraints are satisfied:

$$\sigma_{a|x} \geqslant 0 \quad \forall\, a \in \mathbb{A},\ x \in \mathbb{X}, \tag{2}$$

$$\sum_a \sigma_{a|x} = \sum_a \sigma_{a|x'} \quad \forall\, a \in \mathbb{A},\ x,\ x' \in \mathbb{X}, \tag{3}$$

$$\operatorname{tr}\left\{\sum_a \sigma_{a|x}\right\} = 1 \quad \forall\, x \in \mathbb{X}. \tag{4}$$

A celebrated theorem by Gisin [70] and Hughston, Jozsa, and Wootters [71] (GHJW) shows that post-quantum assemblages in the traditional bipartite setting specified above cannot occur, that is, any common-cause assemblage given by definition 1 admits a quantum realisation (see theorem 9). However, recent research has shown that in multipartite EPR scenarios [65–67] or modified bipartite ones [68] post-quantum steering may arise. In this work we focus on the former case, which we describe below.

2.3. Multipartite EPR scenarios

We focus on multipartite EPR scenarios that consist of multiple Alice-type parties and one Bob. More specifically, consider the case where we have $N+1$ parties: one of them has a system represented by a Hilbert space $\mathcal{H}_B$ (the party called Bob), and each of the remaining parties has two associated sets of classical variables, $\mathbb{X}_k$ and $\mathbb{A}_k$ (with $k \in \{1, \ldots, N\}$ denoting the party), which represent the classical labels of their measurement choices and outcomes. In analogy with the bipartite case, we regard this parties as 'the Alices', and refer to a particular one by Alice$_k$. All parties are distant, perform space-like separated actions on their share of a physical system, and cannot communicate to each other. By performing measurements, then, the Alices update their knowledge on the state preparation of Bob's system. An assemblage $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ in this scenario is then given by the collection of (possibly unnormalised) quantum states $\{\sigma_{a_1 \ldots a_N|x_1 \ldots x_N}\}_{a_k \in \mathbb{A}_k,\, x_k \in \mathbb{X}_k,\, k \in \{1,\ldots,N\}}$, and $\vec{\mathbb{A}}$ and $\vec{\mathbb{X}}$ are short notation for the arrays $(\mathbb{A}_1, \ldots, \mathbb{A}_N)$ and $(\mathbb{X}_1, \ldots, \mathbb{X}_N)$, respectively.

An assemblage $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ is said to admit a *quantum realisation* if there exists a Hilbert space $\mathcal{H}_{A_k}$ for each Alice, a set of measurements $\{\{M^{(k)}_{a|x}\}_{a \in \mathbb{A}_k}\}_{x \in \mathbb{X}_k}$ in $\mathcal{H}_{A_k}$ for each $k \in \{1, \ldots, N\}$, and a quantum state ρ in $\otimes_{k=1}^{N} \mathcal{H}_{A_k} \otimes \mathcal{H}_B$, such that the elements of $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ can be expressed as:

$$\sigma_{a_1 \ldots a_N|x_1 \ldots x_N} = \mathrm{tr}_{A_1 \ldots A_N} \left\{ \otimes_{k=1}^{N} M^{(k)}_{a_k|x_k} \otimes \mathbb{I}_{\mathcal{H}_B} \rho \right\} \quad \forall\, a_k \in \mathbb{A}_k,\ x_k \in \mathbb{X}_k. \tag{5}$$

In these scenarios, then, the most general assemblage that one can mathematically write while complying with the operational constraints of the scenario are given by the following.

Definition 2. Common-cause assemblage, a.k.a. non-signalling assemblage—multipartite EPR scenarios.

Consider a multipartite EPR scenario, where the Alices' measurements and outcomes are labelled by the elements of the sets $\mathbb{X}_k$ and $\mathbb{A}_k$, respectively, for each $k \in \{1, \ldots, N\}$, and Bob's quantum system is represented by the Hilbert space $\mathcal{H}_B$. Then, an assemblage $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ is a common-cause assemblage (equivalently, a non-signalling assemblage) if the following constraints are satisfied:

$$\sigma_{a_1 \ldots a_N|x_1 \ldots x_N} \geqslant 0 \quad \forall\, a_k \in \mathbb{A}_k,\ x_k \in \mathbb{X}_k, \tag{6}$$

$$\sum_{a_k \in \mathbb{A}_k} \sigma_{a_1 \ldots a_k \ldots a_N|x_1 \ldots x_k \ldots x_N} = \sum_{a_k \in \mathbb{A}_k} \sigma_{a_1 \ldots a_k \ldots a_N|x_1 \ldots x'_k \ldots x_N} \tag{7}$$

$$\forall\, a_j \in \mathbb{A}_j \quad \text{and} \quad x_j \in \mathbb{X}_j \quad \text{with } j \neq k,\ x_k,\ x'_k \in \mathbb{X}_k,\ k \in \{1, \ldots, N\},$$

$$\mathrm{tr} \left\{ \sum_{a_1 \ldots a_k \ldots a_N} \sigma_{a_1 \ldots a_k \ldots a_N|x_1 \ldots x_k \ldots x_N} \right\} = 1 \quad \forall\, x_k \in \mathbb{X}_k. \tag{8}$$

Quantum-realisable assemblages and common-cause assemblages are not the only sets of assemblages of interest in the literature and of relevance in this work. One particular set of assemblages that is relevant is that of *almost-quantum* assemblages, which is a set slightly larger than that of quantum-realisable assemblages. The relevance of almost-quantum assemblages is that they are conveniently defined such that testing whether an assemblage admits an almost-quantum realisation amounts to a single instance of a semidefinite program, hence it serves as a convenient numerical tool to outer bound quantum-realisable assemblages and, from an information-theoretical perspective, upper-bound their resourcefulness [60]. In the next subsection, we introduce the almost-quantum assemblages, and the moment matrix formalisation to which they are equivalent, as this makes the link to semidefinite programming.

2.4. Almost-quantum assemblages and moment matrices

The set of almost-quantum assemblages is strongly connected to the set of almost-quantum correlations [27]. Because of this close connection, we will discuss both in this section. As shown in reference [66], this set of assemblages has two equivalent definitions, one of which is defined in terms of moment matrices. Here we opt to introduce their definition in terms of moment matrices, which is linked to semi-definite programming, since this proves convenient for this work. To begin, we will need some useful notation associated with EPR (and Bell) experiments.

Definition 3. Alphabet, properties of words, and equivalence relations between words.

We begin by defining an alphabet, some of the properties of the words you can write with it, and equivalence relations between words. Consider the case where we have N black-box parties in our experiment, which for consistency we refer to as Alices. Let $\mathbb{X}$ and $\mathbb{A}$ be the sets of measurement choices for each Alice, and measurement outcomes for each measurement, respectively[10]. The alphabet $\Upsilon_{N,}$ whose letters are of the form $a|x$, is defined as:

$$\Upsilon_N := \bigcup_{k=1:N} \{a_k|x_k\}_{a_k\in\mathbb{A},\, x_k\in\mathbb{X}}. \tag{9}$$

A **word** is a concatenation of elements of Υ_N. Given two words $\mathbf{v}, \mathbf{w}$, their concatenation $\mathbf{vw}$ yields another word. The word $\mathbf{v}^\dagger$ denotes the word given by the letters of $\mathbf{v}$ written in reverse order. The symbol $\emptyset$ denotes the **empty word**, and has length 0. Finally, the set $\mathcal{S}^*_{\mathrm{E}}$ is the **set of words** of arbitrary length you may write from the letter of Υ_N and the empty word $\emptyset$.

Equivalence relations (a.k.a. **symmetry operations**) among words are the following:

- $\mathbf{vw} = \mathbf{v}\emptyset\mathbf{w}\ \forall\, \mathbf{v}, \mathbf{w} \in \mathcal{S}^*_{\mathrm{E}}$.
- $\mathbf{vv} = \mathbf{v} \forall\, \mathbf{v} \in \Upsilon_N$.
- $a_k|x_k a_{k'}|x_{k'} = a_{k'}|x_{k'} a_k|x_k \forall\, k \neq k' \in \{1, \ldots, N\}$.

A word $\mathbf{v}$ is **null** if, after applying symmetry operations, there is a letter $a_k|x_k$ followed by a letter $a'_k|x_k$ with $a_k \neq a'_k$.

Definition 4. Almost-quantum set of words.

Let $S \subseteq \{1, \ldots, N\}$ be a subset of the black-box parties. Let $\mathbf{a}_S|\mathbf{x}_S$ denote the word given by $\mathbf{a}_S|\mathbf{x}_S = a_{i_1}|x_{i_1} \ldots a_{i_{|S|}}|x_{i_{|S|}}$, where $a_k|x_k \in \Upsilon_N \forall\, k \in S$.

The set of words $\mathcal{S}_{\mathrm{AQ}}$ is given by

$$\mathcal{S}_{\mathrm{AQ}} := \{\emptyset\} \cup \{\mathbf{a}_S|\mathbf{x}_S \mid a_k|x_k \in \Upsilon_N \quad \forall\, k \in S, \quad S \subseteq \{1, \ldots, N\}\}. \tag{10}$$

The following definitions now are particular to correlations in a Bell experiment, or assemblages. When talking about Bell scenarios, we will consider an N-partite scenario with N black-box parties, which for consistency we will call Alices. A correlation in such multipartite Bell scenario will be denoted by $\mathbf{P}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$, where $\vec{\mathbb{A}}$ is an N component vector, where each entry is the set $\mathbb{A}$, and similarly for $\vec{\mathbb{X}}$. For an EPR scenario, we will focus on one with N black-box parties (Alices) and one Bob (whose quantum system has finite dimension d). An assemblage in such EPR scenario will be denoted by $\boldsymbol{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$.

[10] As before, for the purpose of this work and simplicity in the presentation, it is convenient to take the set of measurements to be the same for all Alices, and the set of outcomes to be the same for all measurements.

Definition 5. Bell-scenario almost-quantum moment matrix.

Let $\Gamma^{\mathrm{B}}_{\mathrm{AQ}}$ be a square matrix of size $|\mathcal{S}_{\mathrm{AQ}}| \times |\mathcal{S}_{\mathrm{AQ}}|$, whose entries are complex numbers. Let the rows and columns of $\Gamma^{\mathrm{B}}_{\mathrm{AQ}}$ be labelled by the words in $\mathcal{S}_{\mathrm{AQ}}$. $\Gamma^{\mathrm{B}}_{\mathrm{AQ}}$ is an almost-quantum moment matrix in the Bell scenario iff it satisfies the following properties:

$$\Gamma^{\mathrm{B}}_{\mathrm{AQ}} \geqslant 0, \tag{11}$$

$$\Gamma^{\mathrm{B}}_{\mathrm{AQ}}(\emptyset, \emptyset) = 1, \tag{12}$$

$$\Gamma^{\mathrm{B}}_{\mathrm{AQ}}(\mathbf{v}, \mathbf{w}) = \Gamma^{\mathrm{B}}_{\mathrm{AQ}}(\mathbf{v}', \mathbf{w}') \quad \text{if } \mathbf{v}^{\dagger}\mathbf{w} = \mathbf{v}'^{\dagger}\mathbf{w}', \tag{13}$$

$$\Gamma^{\mathrm{B}}_{\mathrm{AQ}}(\mathbf{v}, \mathbf{w}) = 0 \quad \text{if } \mathbf{v}^{\dagger}\mathbf{w} \text{ is null.} \tag{14}$$

Definition 6. Almost-quantum correlation.

A correlation $\mathbf{P}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ is almost-quantum if there exists an almost-quantum moment matrix $\Gamma^{\mathrm{B}}_{\mathrm{AQ}}$ such that:

$$\Gamma^{\mathrm{B}}_{\mathrm{AQ}}(\emptyset, \mathbf{a}_S|\mathbf{x}_S) = p_S(\mathbf{a}_S|\mathbf{x}_S), \quad \forall\, \mathbf{a}_S|\mathbf{x}_S \in \mathcal{S}_{\mathrm{AQ}}, \tag{15}$$

where $p_S(\mathbf{a}_S|\mathbf{x}_S)$ is the marginal conditional probability distribution of $\mathbf{P}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ for the parties in the set S.

Definition 7. EPR-scenario almost-quantum moment matrix.

Let $\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}$ be a square matrix of size $|\mathcal{S}_{\mathrm{AQ}}| \times |\mathcal{S}_{\mathrm{AQ}}|$, whose entries are $d \times d$ complex matrices. Let the rows and columns of $\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}$ be labelled by the words in $\mathcal{S}_{\mathrm{AQ}}$. $\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}$ is an almost-quantum moment matrix in the EPR scenario iff it satisfies the following properties:

$$\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}} \geqslant 0, \tag{16}$$

$$\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\mathbf{v}, \mathbf{w}) = \Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\mathbf{v}', \mathbf{w}') \quad \text{if } \mathbf{v}^{\dagger}\mathbf{w} = \mathbf{v}'^{\dagger}\mathbf{w}', \tag{17}$$

$$\mathrm{tr}\left\{\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\emptyset, \emptyset)\right\} = 1 \tag{18}$$

$$\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\mathbf{v}, \mathbf{w}) = \mathbf{0} \quad \text{if } \mathbf{v}^{\dagger}\mathbf{w} \text{ is null,} \tag{19}$$

where $\mathbf{0}$ is a $d \times d$ matrix whose entries are all 0.

Definition 8. Almost-quantum assemblage.

An assemblage $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ is almost-quantum if there exists an almost-quantum moment matrix $\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}$ such that:

$$\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\emptyset, \emptyset) = \rho_R, \tag{20}$$

where $\rho_R = \sum_{a_1,\dots a_N \in \mathbb{A}} \sigma_{a_1,\dots,a_N|x_1,\dots,x_N}$ is the reduced state of Bob's system, and

$$\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\emptyset, \mathbf{a}_S|\mathbf{x}_S) = \sigma_S(\mathbf{a}_S|\mathbf{x}_S), \quad \forall\, \mathbf{a}_S|\mathbf{x}_S \in \mathcal{S}_{\mathrm{AQ}}, \tag{21}$$

where $\sigma_S(\mathbf{a}_S|\mathbf{x}_S)$ is the marginal assemblage element of $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ for the parties in the set S:

$$\sigma_S(\mathbf{a}_S|\mathbf{x}_S) = \sum_{a_k \in \mathbb{A}: k \notin S} \sigma_{a_1,\dots,a_N|x_1,\dots,x_N}. \tag{22}$$

Notice that Bob's reduce state ρ_R as well as the marginal assemblage elements $\sigma_S(\mathbf{a}_S|\mathbf{x}_S)$ are well defined since we are working with assemblages $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ that are non-signalling. As a consequence, we have that all almost-quantum assemblages are non-signalling assemblages. However, in general scenarios the converse is not true [65]. It can also be shown that all quantum assemblages are almost-quantum assemblages [65, 66]. Thus, this set is very useful for understanding how to recover the set of a quantum assemblages from a more general set.

3. Characterising assemblages

Let us begin this section by discussing the question of characterising quantum correlations in Bell scenarios. Notice the focus on 'the question' rather than 'the attempted answers'. For simplicity and concreteness, let us base the discussion on the case of correlations in a bipartite Bell scenario. In this case, we have two distant parties—Alice and Bob—which share a physical system and perform measurements on it. The object of interest here is the correlations between their measurement outcomes, captured by the conditional probability distribution $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}} = \{\{p(ab|xy)\}_{a\in\mathbb{A},\, b\in\mathbb{B}}\}_{x\in\mathbb{X},\, y\in\mathbb{Y}}$, where $\mathbb{X}$ (resp. $\mathbb{Y}$) is the set of classical labels of Alice's (resp. Bob's) measurement choices, and $\mathbb{A}$ (resp. $\mathbb{B}$) denotes the set of classical labels of Alice's (resp. Bob's) measurement outcomes[11]. Notice that in this description of a Bell scenario, the party named Bob plays a similar role to that of Alice, unlike in an EPR scenario, and we hope any possible confusion is avoided from context.

In a Bell scenario, then, we want to characterise the set $\mathcal{Q}_{\text{Bell}}$ of correlations $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}}$ that admit a *quantum realisation*. We recall here that $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}}$ admits a quantum realisation[12] if there exists a Hilbert space $\mathcal{H}_A$ associated to Alice, a Hilbert space $\mathcal{H}_B$ associated to Bob, a bipartite quantum system in $\mathcal{H}_A \otimes \mathcal{H}_B$, a density matrix $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B)$, a collection of complete projective measurements $\{\{\Pi^{(\text{A})}_{a|x}\}_{a\in\mathbb{A}}\}_{x\in\mathbb{X}}$ in $\mathcal{H}_A$, and a collection of complete projective measurements $\{\{\Pi^{(\text{B})}_{b|y}\}_{b\in\mathbb{B}}\}_{y\in\mathbb{Y}}$ in $\mathcal{H}_B$, such that

$$p(ab|xy) = \text{tr}\left\{\Pi^{(\text{A})}_{a|x} \otimes \Pi^{(\text{B})}_{b|y}\,\rho\right\}, \quad \forall\, a \in \mathbb{A},\ b \in \mathbb{B}\ x \in \mathbb{X},\ y \in \mathbb{Y}. \tag{23}$$

The question of characterising quantum correlations is then framed as: which principles do we need to impose on $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}}$ so that all and only the ones compatible with them are the quantum ones $\mathcal{Q}_{\text{Bell}}$? Notice then that such principles can then be formulated merely in terms of the probabilities themselves, by imposing constraints on the positive numbers in $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}}$. Probabilities are objects with which we are familiar from probability theory, and the object itself is the same as we would study classically (apart from that, when going beyond classical, it is possible to general different correlations $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}}$). All this is to say that there is a natural way to try and bound the correlations by formulating principles regarding only their statistical predictions[13]. Examples of these principles include *no signalling* (that Alice and Bob cannot use $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}}$ to communicate faster than the speed of light) [17], *non-trivial communication complexity* (that Alice cannot use $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}}$ to send an unbounded amount of information to Bob by transmitting only a single bit of classical communication) [16], and *macroscopic locality* [when the source that prepares Alice and Bob's shared physical system sends them

[11] In principle, different measurement can have different number of outcomes, but for the current discussion we can take these sets to be the same without loss of generality.

[12] To be more precise, this is the definition of quantum realisation in the so-called *tensor product paradigm*.

[13] This approach to characterising correlations is sometimes referred to as *device-independent*. Given how device-independent principles have failed to answer the question fully, it is now conjectured by some that device-independent principles are useful but not sufficient to characterise $\mathcal{Q}_{\text{Bell}}$.

many independent copies of it rather than a single one, then the correlations observed in this macroscopic experiment (under certain assumptions, see reference [23]) are classical even if the single-system correlations $\mathbf{P}_{\mathbb{AB}|\mathbb{XY}}$ are not] [23, 72], among others.

Now let us get back to an EPR scenario, were the Alices have access to classical variables (denoting their measurement choices and outcomes) and Bob has access to a quantum system prepared in a known quantum state. Here, we not only want to characterise the outcome statistics of the Alices' measurements, but also Bob's Hilbert space and his system's quantum state. Hence, the tools and approach from the case of Bell scenarios cannot be directly applied to assemblages. The main challenge here is then what to incorporate in the formulation of the principles such that they capture Bob's quantum nature meaningfully. The types of axioms used to derive quantum theory are quite different from the type of device-independent principles pursued for characterising correlations in Bell scenarios, so the challenge is to find the crucial aspect of each approach that may be relevant for characterising assemblages in EPR scenarios. This comprehensive approach is beyond the scope of this work. Here we take the first steps into approaching the question by asking ourselves: (i) what information should we complement device-independent principles with so as to get a chance at bounding assemblages? And (ii) how useful may such a first step be for tackling the question of interest?

3.1. Bipartite EPR scenarios

Traditional bipartite EPR scenarios are a very special case of EPR scenarios for various reasons, one of them being that, here, the question of 'characterising quantum assemblages' has been fully answered:

Theorem 9. *GHJW theorem* [70, 71].

Consider a bipartite EPR scenario $(\mathbb{A}, \mathbb{X}, \mathcal{H}_B)$*. An assemblage* $\mathbf{\Sigma}_{\mathbb{A}|\mathbb{X}}$ *admits a quantum realisation if and only if it satisfies the no-signalling principle.*

Formally, the mathematical constraints that the no-signalling principle implies are those of equation (3). There are various proofs of this theorem in the literature after the seminal work of Gisin [70] and Hughston, Jozsa, and Wootters [71], using different proof techniques. In the following we include one such proof based on moment matrices, for illustration.

Proof. The structure of the proof is to show that: on the one hand, a non-signalling assemblage in this setting is an almost-quantum assemblage as per definition 7; on the other hand, all almost-quantum assemblages are quantum-realisable assemblages in this setting.

First, recall how almost-quantum assemblages are defined in terms of a moment matrix $\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}$ in the following way:

- The set of words is $\mathcal{S}_{\mathrm{AQ}} := \{\emptyset\} \cup \{a|x\,|a \in \mathbb{A},\ x \in \mathbb{X}\}$ as per definition 4,
- $\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(a|x, a'|x) = \mathbf{0}$ if $a \neq a'$, for all $a \neq a' \in \mathbb{A}$ and $x \in \mathbb{X}$ (equation (19)),
- $\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(a|x, \emptyset) = \Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\emptyset, a|x) = \Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(a|x, a|x)$ for all $a \in \mathbb{A},\ x \in \mathbb{X}$ (equation (17)).

For simplicity, in the rest of this proof we will write $\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}$ as Γ.

Given a non-signalling assemblage there is a moment matrix Γ such that $\Gamma(\emptyset, \emptyset) := \sigma_R$, $\Gamma(\emptyset, a|x) := \sigma_{a|x} =: \Gamma(a|x, \emptyset)$ for all $a \in \mathbb{A},\ x \in \mathbb{X}$. To see that this Γ is a valid moment matrix for the assemblage $\mathbf{\Sigma}_{\mathbb{A}|\mathbb{X}}$ notice that (i) conditions in equations (17)–(21) hold directly by construction, and (ii) condition in equation (16) can be shown by explicitly constructing a positive semi-definite matrix. Take σ_R, and write its purification as $|\psi\rangle = \sum_i \sqrt{\lambda_i}|i\rangle_{\text{aux}}|i\rangle_B$, with $\mathcal{H}_{\text{aux}}$ being an auxiliary Hilbert space used for purification. Define now

$$N_{a|x} := \frac{1}{\sqrt{\sigma_R}}\sigma_{a|x}\frac{1}{\sqrt{\sigma_R}}; \tag{24}$$

and notice that $N_{a|x}$ is a POVM. Thus there is another auxiliary system $\mathcal{H}_{\text{aux}'}$ such that

$$\text{tr}_{\text{aux}}\,\{(N_{a|x}\otimes\mathbb{1})|\psi\rangle\langle\psi|\} \equiv \text{tr}_{\text{aux,aux}}\{\Pi_{a|x}|\Psi\rangle\langle\Psi|'\}, \tag{25}$$

where $|\Psi\rangle = |\psi\rangle\otimes|0\rangle_{\text{aux}'}$. Finally, define a matrix V such that

$$V^\dagger(\emptyset) := |\Psi\rangle; \tag{26}$$

$$V^\dagger(a|x) := \Pi_{a|x}|\Psi\rangle. \tag{27}$$

Here the terms in the brackets denote the particular words that will appear in the final moment matrix, as will become clear now. Define a positive semi-definite matrix $\Sigma = V^\dagger V$. From this matrix we can recover Γ by tracing out the two auxiliary systems with spaces $\mathcal{H}_{\text{aux}}$ and $\mathcal{H}_{\text{aux}'}$, i.e. $\Gamma = \text{tr}_{\text{aux,aux}'}\left\{V^\dagger V\right\}$. Since partial tracing is a CP map and $V^\dagger V \geqslant 0$, thus $\Gamma \geqslant 0$ and every non-signalling assemblage has an almost-quantum moment matrix.

To prove that every almost-quantum moment matrix has a quantum realisation in this bipartite setting, from the moment matrix we can explicitly describe the state and measurements for Alice that give that moment matrix. Given a moment matrix Γ, the sub-matrix $\Gamma(\emptyset)$ is a Gramian matrix such that each element $\Gamma(\mathbf{0})_{ij} := \langle i|\Gamma(\emptyset)|j\rangle = \langle u_j|u_i\rangle$ is the inner product of a set of (possibly sub-normalised) vectors $\{|u_i\rangle \in \mathcal{H}\}_i$ in some Hilbert space $\mathcal{H}_A$, where A is used to denote Alice's system. From this set of vectors, we can construct a quantum state in the space $\mathcal{H}_A\otimes\mathcal{H}_B$:

$$|\psi\rangle = \sum_i |u_i\rangle|i\rangle_B, \tag{28}$$

where $|i\rangle_B$ is a vector in $\mathcal{H}_B$. It can readily be deduced that this gives the correct reduced density matrix, by noting:

$$\text{tr}_A\,\{|\psi\rangle\langle\psi|\} = \sum_k \langle k|_A\left(\sum_i |u_i\rangle|i\rangle_B\sum_j\langle u_j|\langle j|_B\right)|k\rangle_A \tag{29}$$

$$= \sum_{ij}\langle u_j|\sum_k |k\rangle\langle k|_A|u_i\rangle|i\rangle\langle j|_B \tag{30}$$

$$= \sum_{ij}\langle u_j|u_i\rangle|i\rangle\langle j|_B \tag{31}$$

$$= \Gamma(\emptyset). \tag{32}$$

It remains to describe Alice's measurements, which will be projectors $\{\Pi_{a|x}\}_{a,x}$ that project onto $\mathcal{H}_A$. Given a moment matrix Γ, from each sub-matrix $\Gamma_{a|x,a|x}$ we can describe a projective measurement. First note that, as before, each element satisfies $\langle i|\Gamma(a|x,a|x)|j\rangle = \langle u_j^{a|x}|u_i^{a|x}\rangle$, thus with $a|x$ we can associate a set of vectors $\{|u_i^{a|x}\rangle\}_i$ living in the same Hilbert space $\mathcal{H}_A$ as before; this is because both $\Gamma(\emptyset)$ and $\Gamma(a|x,a|x)$ have the same dimension.

Now each of Alice's projectors $\Pi_{a|x}$ just projects onto the span of the set $\{|u_i^{a|x}\rangle\}_i$. First note that since $\Gamma^{\text{EPR}}_{\text{AQ}}(a|x, a'|x) = \mathbf{0}$ if $a \neq a'$, this implies that the span of the set $\{|u_i^{a'|x}\rangle\}_i$ lies in the orthogonal complement of the span of the set $\{|u_i^{a|x}\rangle\}_i$. This property ensures that, for each x, $\Pi_{a|x}\Pi_{a'|x} = \delta_{a,a'}\Pi_{a|x}$. Because $\Gamma(\emptyset, a|x) = \Gamma(a|x, a|x)$, this implies that $\langle u_j|u_i^{a|x}\rangle = \langle u_j^{a|x}|u_i^{a|x}\rangle$, thus both $|u_j\rangle$ and $|u_j^{a|x}\rangle$ have the same inner product with all vectors $|u_i^{a|x}\rangle$ onto which $\Pi_{a|x}$ projects. Therefore, $\Pi_{a|x}|u_i\rangle = |u_i^{a|x}\rangle$ and that $\sum_a \Pi_{a|x} = \mathbb{1}$ since $\sum_a \Gamma(\emptyset, a|x) = \Gamma(\emptyset)$. Finally, given that $\Pi_{a|x}|u_i\rangle = |u_i^{a|x}\rangle$ one can then infer that this complete, projective measurement $\{\Pi_{a|x}\}_a$ recovers the assemblage element $\sigma_{a|x}$ in the following way:

$$\sigma_{a|x} = \text{tr}_A\left\{\Pi_{a|x} \otimes \mathbb{1}|\psi\rangle\langle\psi|\right\} \tag{33}$$

$$= \sum_k \langle k|_A \left(\sum_i \Pi_{a|x}|u_i\rangle|i\rangle_B \sum_j \langle u_j|\langle j|_B\right)|k\rangle_A \tag{34}$$

$$= \sum_{ij} \langle u_j|\sum_k |k\rangle\langle k|_A|u_i^{a|x}\rangle|i\rangle\langle j|_B \tag{35}$$

$$= \sum_{ij} \langle u_j|u_i^{a|x}\rangle|i\rangle\langle j|_B \tag{36}$$

$$= \Gamma(\emptyset, a|x), \tag{37}$$

thus completing the proof. □

From this we see that characterising quantum-realisable assemblages is a fundamentally *multipartite* task. Another way of viewing this is that the non-signalling principle is enough to capture bipartite quantum assemblages.

4. Characterising assemblages by characterising correlations

In this section we discuss how device-independent principles defined to bound correlations in Bell scenarios can be re-purposed to bound the set of assemblages in EPR scenarios. Since there is a close connection between correlations and assemblages, the hope is to leverage our understanding of the former to the case of assemblages.

The first step we take is motivated by the practicalities of the EPR experiment: at the end of the day, for Bob to characterise any assemblage he needs to perform tomography on the state of his system. That is, he will have access to a set of tomographically-complete measurements $\text{TC}_\text{B} := \{\{M_{b|y}\}_b\}_y$ in $\mathcal{H}_B$, and all the information needed to reconstruct the elements of $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ is given by

$$\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}} := \{p(\vec{a}b|\vec{x}y) \equiv \text{tr}\left\{M_{b|y}\sigma_{a_1\ldots a_N|x_1\ldots x_N}\right\} | M_{b|y} \in \text{TC}_\text{B}, \quad \sigma_{a_1\ldots a_N|x_1\ldots x_N} \in \mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}\right\}. \tag{38}$$

For each y, the elements $\{M_{b|y}\}_b$ can be taken to be projectors without loss of generality.

Given a particular but arbitrary principle for correlations in Bell scenarios, one may then apply it to assemblages as follows:

Definition 10. Let $\mathcal{P}$ be a principle for correlations in Bell scenarios, we say that an assemblage $\boldsymbol{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ satisfies $\mathcal{P}$ if the correlation $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ from equation (38)—produced when Bob performs tomographically-complete measurements on the elements of $\boldsymbol{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$—satisfies $\mathcal{P}$.

4.1. Macroscopic classicality considerations

We first illustrate how to use our idea to impose on assemblages the macroscopic locality principle [23], or more precisely, its stronger version called macroscopic non-contextuality [72].

The principle of macroscopic non-contextuality states that a correlation $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ is compatible with a particular classical limit (see references [72, 73]) if and only-if $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ belongs to the so-called almost-quantum set of correlations. So the question is: what is the set of assemblages whose correlations $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ are all and only almost-quantum correlations? We discuss below answers to this question through propositions 11 and 12.

Proposition 11. *Almost-quantum assemblages satisfy macroscopic non-contextuality.*

Proof. An assemblage $\boldsymbol{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ is almost-quantum assemblage if there exist [66, lemma 16]:

- A Hilbert space $\mathcal{H} := \mathcal{K} \otimes \mathcal{H}_B$;
- A state $|\psi\rangle \in \mathcal{H}$;
- A collection of projective measurements $\{\{\Pi^{(i)}_{a_i|x_i}\}_{a_i\in\mathbb{A}}\}_{x_i\in\mathbb{X}}\}$ acting on $\mathcal{K}$, for each party $i = 1, \ldots, N$,

such that

$$\sigma_{a_1\ldots a_N|x_1\ldots x_N} := \mathrm{tr}_{\mathcal{K}}\left\{\Pi^{(1)}_{a_1|x_1}\ldots\Pi^{(N)}_{a_N|x_N}\otimes \mathbb{1}_B|\psi\rangle\langle\psi|\right\}, \tag{39}$$

and

$$\Pi^{(1)}_{a_1|x_1}\ldots\Pi^{(N)}_{a_N|x_N}\otimes \mathbb{1}_B|\psi\rangle = \Pi^{(\pi(1))}_{a_{\pi(1)}|x_{\pi(1)}}\ldots\Pi^{(\pi(N))}_{a_{\pi(N)}|x_{\pi(N)}}\otimes \mathbb{1}_B|\psi\rangle, \tag{40}$$

for all permutations π of the N Alices.

Let $\{\Pi_{b|y}\}_{b\in\mathbb{B},y\in\mathbb{Y}}$ be a collection of projective measurements for Bob (not necessarily a tomographically complete one), that he could perform over the elements of $\Sigma_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$.

The correlations $\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ arising from these measurements then satisfy the following property: there exist

- A Hilbert space $\mathcal{H}$;
- A state $|\psi\rangle \in \mathcal{H}$;
- A collection of projective measurements $\{\Pi^{(i)}_{a_i|x_i}\}_{a_i\in\mathbb{A},x_i\in\mathbb{X},i=1,\ldots,N} \cup \{\Pi_{b|y}\}_{b\in\mathbb{B},y\in\mathbb{Y}}$ acting on $\mathcal{H}$,

such that

$$p(\vec{a}b|\vec{x}y) := \mathrm{tr}\left\{\Pi_{b|y}\sigma_{a_1\ldots a_N}\right\} = \mathrm{tr}\left\{\Pi^{(1)}_{a_1|x_1}\ldots\Pi^{(N)}_{a_N|x_N}\otimes\Pi_{b|y}|\psi\rangle\langle\psi|\right\}. \tag{41}$$

Now, since Bob's measurements commute with all of Alices' measurements, then equation (41) provides an almost-quantum realisation of the correlations $\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$.

The last step is to recall that almost-quantum correlations satisfy the macroscopic noncontextuality principle [72]. Hence, almost-quantum assemblages also satisfy the principle in the sense of definition 10. □

We have therefore shown that almost-quantum assemblages satisfy the macroscopic noncontextuality principle. Now, is the converse true? That is, is this an 'if and only if' result? The answer to this is not straightforward, and requires one to leverage some physical characteristics of an EPR experiment, as we discuss next.

For the argument, notice that there exist assemblages $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ such that:

- The correlations $\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ that may arise when Bob performs any measurements on his system (including but not restricted to tomographically-complete ones) admit a quantum realisation,
- The assemblage $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ is not almost-quantum.

As we will discuss in detail later, in [65], for instance, such an example of an assemblage was given. One way to understand this it that the quantum realisation of $\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ may require a quantum system in Bob's wing that has larger dimension than that of $\mathcal{H}_B$ (Bob's Hilbert space in the EPR scenario), hence explaining why the correlations admit a quantum realisation but the assemblage does not. From this fact it follows that there exist assemblages that generate almost-quantum correlations but are themselves not almost-quantum.

The examples of such assemblages might come across as evidence that assemblages beyond the almost-quantum set satisfy the macroscopic non-contextuality principle. And this indeed highlights a deficiency of definition 10: one needs additional information beyond the correlations $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ to fully flesh-out the structure of assemblages in EPR scenarios. See section 4.3 for a more comprehensive device-dependent discussion of the claim stated here.

4.2. The need for device-dependent principles

The task that we set out to explore is that of characterising quantum assemblages, that is, a collection of quantum states. Hence, it comes as no surprise that one may require principles that go beyond the device-independent ones. In the last section, however, we saw how one particular device-dependent way in which device-independent principles may be applied to assemblages failed at answering our question. Here we briefly discuss other device-dependent considerations one may leverage.

One of the captivating properties of post-quantum EPR inference is that it is a new post-quantum phenomenon in its own right, which is not merely implied by post-quantumness in Bell scenarios. However, when it comes to characterising assemblages solely from device-independent principles this becomes a roadblock. More precisely, consider an assemblage $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$, and a set of measurements $\{\{N_{b|y}\}_{b\in\mathbb{B}}\}_{y\in\mathbb{Y}}\}$ on $\mathcal{H}_B$ for Bob (not necessarily tomographically complete), where $\mathbb{Y}$ is the set of classical labels for the measurement choices, and $\mathbb{B}$ the set of classical labels for the measurements' outcomes. One may compute the correlations

$$\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}} := \Big\{ p(\vec{a}b|\vec{x}y) \equiv \operatorname{tr}\left\{N_{b|y}\sigma_{a_1\ldots a_N|x_1\ldots x_N}\right\} \,|\, y\in\mathbb{Y},\; b\in\mathbb{B},\; \sigma_{a_1\ldots a_N|x_1\ldots x_N}\in\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}} \Big\}. \tag{42}$$

Notice that the operational constraints in the scenario where the Alices and Bob generate the correlations $\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ are consistent with those of a Bell scenario, and hence we can think of $\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ as being 'correlations in a Bell experiment'. Now, references [65, 67] show that there exist post-quantum assemblages $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ that can only generate quantum correlations $\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ in such a Bell scenario. Therefore, any principle imposed on $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ which only looks at the possible correlations $\mathbf{P}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ that it may generate *without involving any information on how these correlations were generated*, will immediately be satisfied by some post-quantum assemblages, including assemblages with no almost-quantum realisation.

In the previous subsection we discussed a particular way in which device-dependent considerations can be brought into the picture to bound EPR assemblages. This approach, presented in definition 10, complements these device-independent principles with the particular information on the dimension of Bob's system $\mathcal{H}_B$ as well as the particular characterisation of his measurement device which is also required to perform tomographically-complete measurements. What definition 10 does not explicitly state—and which some may feel inclined to include it in retrospect—is that when thinking of $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ as a 'correlation in a Bell scenario' one can moreover require that the realisation of $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$—be it quantum, almost-quantum, etc—should have Bob's Hilbert space factorised from whatever object represents Alices' systems, and his quantum actions should act locally on $\mathcal{H}_B$. In the next subsection we show how to apply this reasoning when asking how the macroscopic non-contextuality principle constrains EPR assemblages.

4.3. A stronger device-dependent version of macroscopic non-contextuality

We saw in section 4.1 how the minimalistic application of the macroscopic non-contextuality principle to assemblages via definition 10 fails at constraining the set of assemblages in a satisfactory way. The question we discuss here is how we can complement definition 10 with device-dependent requirements so that a more meaningful set of assemblages is singled out by the macroscopic non-contextuality principle.

The first consideration that stems from the EPR scenario is that Bob's Hilbert space is known and specified ($\mathcal{H}_B$), hence one can argue that the required specific (such as quantum or almost-quantum) realisation of $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ does happen in a factorised Hilbert space $\mathcal{H} := \mathcal{K} \otimes \mathcal{H}_B$. In the same way, a natural requirement is that Bob's measurement operators that realise the correlations only act on $\mathcal{H}_B$.

A second consideration is about the Alices. Given the previous discussion, if one requires the correlations $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ to be realised in $\mathcal{H} := \mathcal{K} \otimes \mathcal{H}_B$ one may also take a step forward demanding that the operations implemented by the Alices when realising the correlations only act on $\mathcal{K}$. While this is a more questionable assumption, it is worth considering, although it is not required for the proof of our following claim.

Proposition 12. *Consider an EPR scenario* $(\vec{\mathbb{X}}, \vec{\mathbb{A}}, \mathcal{H}_B)$*, and an assemblage* $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ *representing the knowledge on the state of Bob's system. Let* $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ *be the correlation associated to* $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ *when Bob performs tomographically-complete measurements on the latter. Assume that:*

- $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ *is an almost-quantum correlation, that is, that* $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ *satisfies the macroscopic non-contextuality principle as per definition 10,*
- *The almost-quantum realisation of* $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ *is achieved in a factorised Hilbert space* $\mathcal{H} := \mathcal{K} \otimes \mathcal{H}_B$ *by measurements* $\{\Pi_{b|y}\}_{b\in\mathbb{B},y\in\mathbb{Y}}$ *for Bob with the form* $\Pi_{b|y} = \mathbb{1}_{\mathcal{K}} \otimes \Pi^{\mathrm{TC}}_{b|y}$, $\forall\, b \in \mathbb{B},\ y \in \mathbb{Y}$.

Then, $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ *is an almost-quantum assemblage.*

Proof. Consider the almost quantum certificate $\Gamma^{\mathrm{B}}_{\mathrm{AQ}}$ for the correlations $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$. Denoting by $\mathcal{S}_{\mathrm{AQ}}$ the almost quantum set of words for all black-box parties' labels, it is the case that $\Gamma^{\mathrm{B}}_{\mathrm{AQ}} \in M_{|\mathcal{S}_{\mathrm{AQ}}|\cdot|\mathbb{B}|\cdot|\mathbb{Y}|}(\mathbb{R})$, where $M_n(\mathbb{R})$ is the space of $n \times n$ matrices with real elements. Furthermore, there is an isomorphism between this space and $M_{|\mathcal{S}_{\mathrm{AQ}}|}(M_{|\mathbb{B}|\cdot|\mathbb{Y}|}(\mathbb{R}))$, i.e., the space of $|\mathcal{S}_{\mathrm{AQ}}| \times |\mathcal{S}_{\mathrm{AQ}}|$ matrices with elements from $M_{|\mathbb{B}|\cdot|\mathbb{Y}|}(\mathbb{R})$. From this, we can represent $\Gamma^{\mathrm{B}}_{\mathrm{AQ}}$ as

$$\Gamma^{\mathrm{B}}_{\mathrm{AQ}} = \sum_{\mathbf{v},\mathbf{w}\in\mathcal{S}_{\mathrm{AQ}}} |\mathbf{v}\rangle\langle\mathbf{w}| \otimes \Lambda_{\mathbf{v},\mathbf{w}}, \tag{43}$$

where each $\Lambda_{\mathbf{v},\mathbf{w}} \in M_{|\mathbb{B}|\cdot|\mathbb{Y}|}(\mathbb{R})$ is a submatrix of $\Gamma^{\mathrm{B}}_{\mathrm{AQ}}$. Notice, moreover, that when $\mathbf{w}^\dagger\mathbf{v} \in \mathcal{S}_{\mathrm{AQ}}$, $\Lambda_{\mathbf{v},\mathbf{w}}$ contains the probabilities and correlations of Bob's tomographic measurements all conditioned to the labels $\mathbf{w}^\dagger\mathbf{v}$ of the black-box parties.

Because we know that Bob's measurements act only on $\mathcal{H}_B$ in the factorized Hilbert space, and because they are tomographically complete, there is an isomorphism between the space $M_{|\mathbb{B}|\cdot|\mathbb{Y}|}(\mathbb{R})$ and Bob's Hilbert space $\mathcal{H}_B$, i.e., there is a bijective, positive map $\mathcal{T}(\cdot)$ such that:

$$\mathcal{T}(\Lambda_{\emptyset,\emptyset}) = \sigma_R; \tag{44}$$

$$\mathcal{T}(\Lambda_{\emptyset,\mathbf{v}}) = \mathcal{T}(\Lambda_{\mathbf{v},\mathbf{v}}) = \sigma_{\mathbf{v}}. \tag{45}$$

These elements (σ_R and $\sigma_{\mathbf{v}}$) are positive semidefinite by definition, since the corresponding $\Lambda_{\emptyset,\emptyset}$ and $\Lambda_{\mathbf{v},\mathbf{v}}$ are positive semidefinite themselves (the latter follows from Sylverster's criterion, since $\Lambda_{\emptyset,\emptyset}$ and $\Lambda_{\mathbf{v},\mathbf{v}}$ are principal submatrices of $\Gamma^{\mathrm{B}}_{\mathrm{AQ}}$ and hence are also positive semidefinite). Notice that, if the Hilbert space were not factorized, or if Bob's measurements would not act on $\mathcal{H}_B$ only, the tomographic reconstruction could yield non-physical states. In other words, the isomorphism is in fact between a subset of $M_{|\mathbb{B}|\cdot|\mathbb{Y}|}(\mathbb{R})$, whose entries are correlations arising from Bob's measurements acting on $\mathcal{H}_B$, and $\mathcal{H}_B$ itself.

From the linearity of $\mathcal{T}(\cdot)$, we can define

$$\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}} \coloneqq (\mathbb{1} \otimes \mathcal{T})[\Gamma^{\mathrm{B}}_{\mathrm{AQ}}] = \sum_{\mathbf{v},\mathbf{w}\in\mathcal{S}_{\mathrm{AQ}}} |\mathbf{v}\rangle\langle\mathbf{w}| \otimes \mathcal{T}(\Lambda_{\mathbf{v},\mathbf{w}}). \tag{46}$$

This matrix has labels and columns completely specified by the words in $\mathcal{S}_{\mathrm{AQ}}$, each entry being a matrix in $\mathcal{H}_B$. It also has the following properties:

$$\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\mathbf{v},\mathbf{w}) = \Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\mathbf{v}',\mathbf{w}') \quad \text{if } \mathbf{v}^\dagger\mathbf{w} = \mathbf{v}'^\dagger\mathbf{w}', \tag{47}$$

$$\mathrm{tr}\left\{\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\emptyset,\emptyset)\right\} = 1, \tag{48}$$

$$\Gamma^{\mathrm{EPR}}_{\mathrm{AQ}}(\mathbf{v},\mathbf{w}) = \mathbf{0} \quad \text{if } \mathbf{v}^\dagger\mathbf{w} \,\text{is null}, \tag{49}$$

where $\mathbf{0}$ is a $d \times d$ matrix with null elements, d being the dimension of $\mathcal{H}_B$. The first property follows from the fact that $\Lambda_{\mathbf{v},\mathbf{w}} = \Lambda_{\mathbf{v}',\mathbf{w}'}$ when $\mathbf{v}^\dagger\mathbf{w} = \mathbf{v}'^\dagger\mathbf{w}'$ and this equality is preserved by the isomorphism. The second comes from the constraint that $\Lambda_{\emptyset,\emptyset}$ has as entries $\mathrm{tr}\left\{(\mathbb{1}_{\mathcal{K}} \otimes \Pi^{\mathrm{TC}}_{b|y})|\psi\rangle\langle\psi|\right\}$, the tomographic data that is mapped into Bob's partial state $\mathrm{tr}_{\mathcal{K}}\left\{|\psi\rangle\langle\psi|\right\}$. The third arises from $\Lambda_{\mathbf{v},\mathbf{w}} = \mathbf{0}$ whenever $\mathbf{v}^\dagger\mathbf{w}$ is null and the linearity of the isomorphism.

To finally show that $\Gamma^{\text{EPR}}_{\text{AQ}} \geqslant 0$, notice that there is a state $|\psi\rangle \in \mathcal{K} \otimes \mathcal{H}_B$ and projectors $\{\Pi^{(i)}_{a_i|x_i}\}_{a\in\mathbb{A},x\in\mathbb{X},i=1,\dots,N}$ acting on this Hilbert space, such that

$$\mathcal{T}(\Lambda_{\mathbf{v},\mathbf{w}}) = \text{tr}_{\mathcal{K}}\left\{\left(\prod_{(a_i|x_i)\in\mathbf{v}} \Pi^{(i)}_{a_i|x_i}\right)^{\dagger} \prod_{(a'_j|x'_j)\in\mathbf{w}} \Pi^{(j)}_{a'_j|x'_j}|\psi\rangle\langle\psi|\right\}. \tag{50}$$

The entry $\Gamma^{\text{EPR}}_{\text{AQ}}(\mathbf{v}_l, \mathbf{w}_m)$, where l, m are labels of the rows and columns in $\mathcal{H}_B$, can be written as

$$\Gamma^{\text{EPR}}_{\text{AQ}}(\mathbf{v}_l, \mathbf{w}_m) = \sum_{|k\rangle\in\mathcal{K}} \langle k|\langle l| \prod_{(a'_j|x'_j)\in\mathbf{w}} \Pi^{(j)}_{a'_j|x'_j}|\psi\rangle\langle\psi| \left(\prod_{(a_i|x_i)\in\mathbf{v}} \Pi^{(i)}_{a_i|x_i}\right)^{\dagger} |k\rangle|m\rangle, \tag{51}$$

or reorganizing and including an identity $\sum_{|k'\rangle\in\mathcal{K}}|k'\rangle\langle k'|$,

$$\Gamma^{\text{EPR}}_{\text{AQ}}(\mathbf{v}_l, \mathbf{w}_m) = \sum_{|k\rangle\in\mathcal{K}} \langle\psi| \left(\prod_{(a_i|x_i)\in\mathbf{v}} \Pi^{(i)}_{a_i|x_i}\right)^{\dagger} |k\rangle|m\rangle\langle k| \cdot \sum_{|k'\rangle\in\mathcal{K}} \langle k'|\langle l| \times \prod_{(a'_j|x'_j)\in\mathbf{w}} \Pi^{(j)}_{a'_j|x'_j}|\psi\rangle|k'\rangle, \tag{52}$$

which can be rewritten as

$$\Gamma^{\text{EPR}}_{\text{AQ}}(\mathbf{v}_l, \mathbf{w}_m) = \langle \mathbf{v}_l|\mathbf{w}_m\rangle. \tag{53}$$

It is thus possible to define a matrix V whose columns are the vectors $|\mathbf{v}_l\rangle$, $\forall\, \mathbf{v} \in \mathcal{S}_{\text{AQ}}$ and $l = 1, \dots, \dim\, \mathcal{H}_B$, such that

$$\Gamma^{\text{EPR}}_{\text{AQ}} = V^{\dagger} V, \tag{54}$$

and therefore, positive semidefinite. □

5. Discussion

In this work we considered the question of how to characterise a natural set of assemblages from basic physical principles in EPR scenarios—a question that has been vastly explored in the device-independent framework for correlations in Bell scenarios. The hope is that if a wise set of principles is identified, the natural set of assemblages that emerges will be that of the quantum-realisable ones, gaining a valuable physical intuition behind quantum assemblages. For the case of correlations in Bell scenarios this question still remains open, and there is no reason *a priori* to expect that it becomes less difficult when asked for assemblages in EPR scenarios.

In this work we explored a particular avenue towards characterising EPR assemblages: the idea is to map an assemblage $\mathbf{\Sigma}_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ into a conditional probability distribution $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ in a particular Bell scenario by letting the characterised party (Bob) perform a set of tomographically-complete measurements on their quantum system. We hence explore how the assemblages

$\Sigma_{\vec{\mathbb{A}}|\vec{\mathbb{X}}}$ are constrained by imposing device-independent principles on their corresponding correlations $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$. The motivation is that, since Bob performs tomographically complete measurements, all the information about the assemblage is encoded in $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$, hence one would expect that exploring the characterisations of the latter would be sufficient. However, we find that the tensor-product structure of the joint Hilbert space underpinning the EPR experiment—in particular, that Bob's Hilbert space is tensored to that of the Alices—is not a property one may capture by merely studying the correlations $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$, and hence the need for truly device-dependent principles for characterising EPR assemblages emerges.

To explore the extent to which our particular avenue may characterise EPR assemblages from their correlations, we focused on the case of correlations $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ that satisfy the macroscopic non-contextuality principle. Here we showed that almost-quantum assemblages satisfy the macroscopic non-contextuality principle—in other words, the correlations $\mathbf{P}^{\mathbf{T}}_{\vec{\mathbb{A}}\mathbb{B}|\vec{\mathbb{X}}\mathbb{Y}}$ that arise from almost-quantum assemblages satisfy macroscopic non-contextuality. The converse statement—that assemblages which satisfy macroscopic non-contextuality are almost-quantum—needs to be formalised in more detail since, as we argued before, the tensor-product structure of the Hilbert space into the product of the parties' needs to be assumed. Grounded in this additional assumption, we further showed that an assemblage satisfies the macroscopic non-contextuality principle only if it is an almost-quantum assemblage.

Our proof techniques involve EPR inferential tools that are still novel, and we hope our work highlights their usefulness for future research in the field. Looking ahead, we have taken the first steps toward tackling the characterisation of EPR inference, and this has open—as is usually the case in foundational research—the door to a plethora of questions, a couple of which are the following. First, one may wonder whether stronger constraints on EPR assemblages can be found by relating them to the objects of study in other non-classicality experiments—for example, correlations in contextuality experiments or network non-classicality rather than the Bell scenarios considered here. Alternatively, one can explore the more challenging avenue of characterising assemblages via device-dependent principles formulated directly for EPR scenarios, without having better-known phenomena (such as Bell non-classicality or contextuality) mediate between them. In this case, toy theories that include quantum systems—such as Witworld [74]—may serve as a test-bed for the strength of the proposed principles.

Acknowledgments

We thank Valerio Scarani for fruitful discussions during the Vienna Quantum Foundations Conference 2020 (hosted in September 2021). VPR and ABS acknowledge support by the Foundation for Polish Science (IRAP project, ICTQT, Contract No. 2018/MAB/5, co-financed by EU within Smart Growth Operational Programme). MJH and ABS acknowledge the FQXi large Grant 'The Emergence of Agents from Causal Order' (FQXi FFF Grant Number FQXi-RFP-1803B). The diagrams within this manuscript were prepared using TikZit.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Ana Belén Sainz https://orcid.org/0000-0003-3123-8436

References

[1] Howard M, Wallman J, Veitch V and Emerson J 2014 Contextuality supplies the 'magic' for quantum computation *Nature* **510** 351–5
[2] Raussendorf R 2013 Contextuality in measurement-based quantum computation *Phys. Rev.* A **88** 022322
[3] Anders J and Browne D E 2009 Computational power of correlations *Phys. Rev. Lett.* **102** 050502
[4] Hoban M J and Browne D E 2011 Stronger quantum correlations with loophole-free postselection *Phys. Rev. Lett.* **107** 120402
[5] Branciard C, Cavalcanti E G, Walborn S P, Scarani V and Wiseman H M 2012 One-sided device-independent quantum key distribution: security, feasibility, and the connection with steering *Phys. Rev.* A **85** 010301
[6] Acín A, Brunner N, Gisin N, Massar S, Pironio S and Scarani V 2007 Device-independent security of quantum cryptography against collective attacks *Phys. Rev. Lett.* **98** 230501
[7] Vazirani U and Vidick T 2019 Fully device independent quantum key distribution *Commun. ACM* **62** 133
[8] Pironio S *et al* 2010 Random numbers certified by Bell's theorem *Nature* **464** 1021–4
[9] Coudron M and Henry Y 2014 Infinite randomness expansion with a constant number of devices *Proc. 46th Annual ACM Symp. Theory of Computing, STOC '14* (New York: Association for Computing Machinery) pp 427–36
[10] Colbeck R and Renner R 2012 Free randomness can be amplified *Nat. Phys.* **8** 450–3
[11] Reichardt B W, Unger F and Vazirani U 2013 Classical command of quantum systems *Nature* **496** 456
[12] Wolfe E, Schmid D, Sainz A B, Kunjwal R and Spekkens R W 2020 Quantifying Bell: the resource theory of nonclassicality of common-cause boxes *Quantum* **4** 280
[13] Bell J S 1964 On the Einstein Podolsky Rosen paradox *Phys. Phys. Fiz.* **1** 195–200
[14] Cirel'son B S 1980 Quantum generalizations of Bell's inequality *Lett. Math. Phys.* **4** 93–100
[15] Clauser J F, Horne M A, Shimony A and Holt R A 1969 Proposed experiment to test local hidden-variable theories *Phys. Rev. Lett.* **23** 880–4
[16] Van Dam W 2013 Implausible consequences of superstrong nonlocality *Nat. Comput.* **12** 9–12
[17] Popescu S and Rohrlich D 1994 Quantum nonlocality as an axiom *Found. Phys.* **24** 379–85
[18] Brassard G, Buhrman H, Linden N, Méthot A A, Tapp A and Unger F 2006 Limit on nonlocality in any world in which communication complexity is not trivial *Phys. Rev. Lett.* **96** 250401
[19] Navascués M, Pironio S and Acín A 2007 Bounding the set of quantum correlations *Phys. Rev. Lett.* **98** 010401
[20] Navascués M, Pironio S and Acín A 2008 A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations *New J. Phys.* **10** 073013
[21] Pironio S, Navascués M and Acín A 2010 Convergent relaxations of polynomial optimization problems with noncommuting variables *SIAM J. Optim.* **20** 2157–80
[22] Pawłowski M, Paterek T, Kaszlikowski D, Scarani V, Winter A and Żukowski M 2009 Information causality as a physical principle *Nature* **461** 1101–4
[23] Navascués M and Wunderlich H 2010 A glance beyond the quantum model *Proc. R. Soc.* A **466** 881–90
[24] Fritz T, Sainz A B, Augusiak R, Brask J B, Chaves R, Leverrier A and Acín A 2013 Local orthogonality as a multipartite principle for quantum correlations *Nat. Commun.* **4** 2263
[25] Sainz A B, Fritz T, Augusiak R, Bohr Brask J, Chaves R, Leverrier A and Acín A 2014 Exploring the local orthogonality principle *Phys. Rev.* A **89** 032117
[26] Contreras-Tejada P, Scarpa G, Kubicki A M, Brandenburger A and La Mura P 2021 Observers of quantum systems cannot agree to disagree *Nat. Commun.* **12** 7021
[27] Navascués M, Guryanova Y, Hoban M J and Acín A 2015 Almost quantum correlations *Nat. Commun.* **6** 6288
[28] Hardy L 2001 Quantum theory from five reasonable axioms (arXiv:quant-ph/0101012)

[29] Clifton R, Bub J and Halvorson H 2003 Characterizing quantum theory in terms of information-theoretic constraints *Found. Phys.* **33** 1561–91
[30] Goyal P 2008 Information-geometric reconstruction of quantum theory *Phys. Rev.* A **78** 052120
[31] Chiribella G, D'Ariano G M and Perinotti P 2010 Probabilistic theories with purification *Phys. Rev.* A **81** 062348
[32] Hardy L 2011 Reformulating and reconstructing quantum theory (arXiv:1104.2066)
[33] Masanes L and Müller M P 2011 A derivation of quantum theory from physical requirements *New J. Phys.* **13** 063001
[34] Chiribella G, D'Ariano G M and Perinotti P 2011 Informational derivation of quantum theory *Phys. Rev.* A **84** 012311
[35] Dakić B and Brukner Č 2011 *Quantum Theory and Beyond: Is Entanglement Special?* (Cambridge: Cambridge University Press) pp 365–92
[36] Hardy L 2012 The operator tensor formulation of quantum theory *Phil. Trans. R. Soc.* A **370** 3385–417
[37] Hardy L 2013 A formalism-local framework for general probabilistic theories, including quantum theory *Math. Struct. Comput. Sci.* **23** 399–440
[38] Budiyono A and Rohrlich D 2017 Quantum mechanics as classical statistical mechanics with an ontic extension and an epistemic restriction *Nat. Commun.* **8** 1306
[39] Höhn P A 2017 Quantum theory from rules on information acquisition *Entropy* **19** 98
[40] Höhn P A 2017 Toolbox for reconstructing quantum theory from rules on information acquisition *Quantum* **1** 38
[41] Selby J H, Scandolo C M and Coecke B 2021 Reconstructing quantum theory from diagrammatic postulates *Quantum* **5** 445
[42] Wilce A 2018 A royal road to quantum theory (or thereabouts) *Entropy* **20** 227
[43] van de Wetering J 2019 An effect-theoretic reconstruction of quantum theory *Compositionality* **1** 1
[44] Nakahira K 2020 Derivation of quantum theory with superselection rules *Phys. Rev.* A **101** 022104
[45] Tull S 2020 A categorical reconstruction of quantum theory *Log. Methods Comput. Sci.* **16** 1–39
[46] Barrett J 2007 Information processing in generalized probabilistic theories *Phys. Rev.* A **75** 032304
[47] Coecke B and Kissinger A 2017 *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning* (Cambridge: Cambridge University Press)
[48] Barrett J, de Beaudrap N, Hoban M J and Lee C M 2019 The computational landscape of general physical theories *npj Quantum Inf.* **5** 41
[49] Oreshkov O, Costa F and Brukner Č 2012 Quantum correlations with no causal order *Nat. Commun.* **3** 1092
[50] Barnum H, Beigi S, Boixo S, Elliott M B and Wehner S 2010 Local quantum measurement and no-signaling imply quantum correlations *Phys. Rev. Lett.* **104** 140401
[51] Acín A, Augusiak R, Cavalcanti D, Hadley C, Korbicz J K, Lewenstein M, Masanes L and Piani M 2010 Unified framework for correlations in terms of local quantum observables *Phys. Rev. Lett.* **104** 140404
[52] Geller J and Piani M 2014 Quantifying non-classical and beyond-quantum correlations in the unified operator formalism *J. Phys. A: Math. Theor.* **47** 424030
[53] Masanes L, Galley T D and Müller M P 2019 The measurement postulates of quantum mechanics are operationally redundant *Nat. Commun.* **10** 1361
[54] Schrödinger E 1936 Probability relations between separated systems *Math. Proc. Camb. Phil. Soc.* **32** 446–52
[55] Wiseman H M, Jones S J and Doherty A C 2007 Steering, entanglement, nonlocality, and the Einstein–Podolsky–Rosen paradox *Phys. Rev. Lett.* **98** 140402
[56] Einstein A, Podolsky B and Rosen N 1935 Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **47** 777
[57] Pusey M F 2013 Negativity and steering: a stronger Peres conjecture *Phys. Rev.* A **88** 032313
[58] Law Y Z, Thinh L P, Bancal J-D and Scarani V 2014 Quantum randomness extraction for various levels of characterization of the devices *J. Phys. A: Math. Theor.* **47** 424028
[59] Passaro E, Cavalcanti D, Skrzypczyk P and Acín A 2015 Optimal randomness certification in the quantum steering and prepare-and-measure scenarios *New J. Phys.* **17** 113010
[60] Cavalcanti D and Skrzypczyk P 2016 Quantum steering: a review with focus on semidefinite programming *Rep. Prog. Phys.* **80** 024001
[61] Skrzypczyk P, Navascués M and Cavalcanti D 2014 Quantifying Einstein–Podolsky–Rosen steering *Phys. Rev. Lett.* **112** 180404

[62] Cavalcanti D, Skrzypczyk P, Aguilar G H, Nery R V, Souto Ribeiro P H and Walborn S P 2015 Detection of entanglement in asymmetric quantum networks and multipartite quantum steering *Nat. Commun.* **6** 7941
[63] Zjawin B, Schmid D, Hoban M J and Sainz A B 2021 Quantifying EPR: the resource theory of nonclassicality of common-cause assemblages (arXiv:2111.10244)
[64] Wood C J and Spekkens R W 2015 The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning *New J. Phys.* **17** 033002
[65] Sainz A B, Brunner N, Cavalcanti D, Skrzypczyk P and Vértesi T 2015 Postquantum steering *Phys. Rev. Lett.* **115** 190403
[66] Hoban M J and Sainz A B 2018 A channel-based framework for steering, non-locality and beyond *New J. Phys.* **20** 053048
[67] Sainz A B, Aolita L, Piani M, Hoban M J and Skrzypczyk P 2018 A formalism for steering with local quantum measurements *New J. Phys.* **20** 083040
[68] Sainz A B, Hoban M J, Skrzypczyk P and Aolita L 2020 Bipartite postquantum steering in generalized scenarios *Phys. Rev. Lett.* **125** 050404
[69] Cavalcanti P J, Selby J H and Sainz A B A generalised probabilistic theory that realises all non-signalling assemblages and boxes (in preparation).
[70] Gisin N 1989 Stochastic quantum dynamics and relativity *Helv. Phys. Acta* **62** 363–71
[71] Hughston L P, Jozsa R and Wootters W K 1993 A complete classification of quantum ensembles having a given density matrix *Phys. Lett.* A **183** 14–8
[72] Henson J and Sainz A B 2015 Macroscopic noncontextuality as a principle for almost-quantum correlations *Phys. Rev.* A **91** 042114
[73] Rossi V P, Hoban M J and Sainz A B Macroscopic limits of steering assemblages (in preparation).
[74] Cavalcanti P J, Selby J H, Sikora J, Galley T D and Sainz A B 2021 Witworld: a generalised probabilistic theory featuring post-quantum steering (arXiv:2102.06581)

Linear Program for Testing Nonclassicality and an Open-Source Implementation

John H. Selby,[1,*] Elie Wolfe,[2,†] David Schmid,[1,‡] Ana Belén Sainz,[1,§] and Vinicius P. Rossi[1,∥]
[1]*International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-309 Gdańsk, Poland*
[2]*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario N2L 2Y5, Canada*

(Received 16 November 2022; revised 12 November 2023; accepted 22 December 2023; published 30 January 2024)

A well-motivated method for demonstrating that an experiment resists any classical explanation is to show that its statistics violate generalized noncontextuality. We here formulate this problem as a linear program and provide an open-source implementation of it which tests whether or not any given prepare-measure experiment is classically explainable in this sense. The input to the program is simply an arbitrary set of quantum states and an arbitrary set of quantum effects; the program then determines if the Born rule statistics generated by all pairs of these can be explained by a classical (noncontextual) model. If a classical model exists, it provides an explicit model. If it does not, then it computes the minimal amount of noise that must be added such that a model does exist, and then provides this model. We generalize all these results to arbitrary generalized probabilistic theories (and accessible fragments thereof) as well; indeed, our linear program is a test of simplex embeddability as introduced in Schmid *et al.* [PRX Quantum **2**, 010331 (2021).] and generalized in Selby *et al.* [Phys. Rev. A **107**, 062203 (2023).].

DOI: 10.1103/PhysRevLett.132.050202

A rigorous method for demonstrating that a theory or a set of data resists any classical explanation is to prove that it cannot be reproduced in any generalized noncontextual model [1]. Generalized noncontextuality was first introduced as an improvement on Kochen-Specker's assumption of noncontextuality [2], making it more operationally accessible and providing stronger motivations for it, as a form of Leibniz's principle [3]. Since its inception, the list of motivations for taking it as one's notion of classicality has grown greatly. Notably, the existence of a generalized-noncontextual ontological model for an operational theory coincides with two independent notions of classicality: one that arises in the study of generalized probabilistic theories [4–6], and another that arises in quantum optics [4,6,7]. Generalized noncontextuality has been used as an indicator of classicality in the quantum Darwinist program [8], and any sufficiently noisy theory satisfies generalized noncontextuality [9,10]. Furthermore, violations of local causality [11], violations of Kochen-Specker noncontextuality [9,12], and some observations of anomalous weak values [13,14], are also instances of generalized contextuality. Finally, generalized contextuality is a resource for information processing [15–19], computation [20], state discrimination [21–24], cloning [25], and metrology [26]. Herein, we use the term *noncontextuality* to refer to the concept of generalized noncontextuality.

How, then, does one determine in practice whether a given theory or a given set of experimental data admits a classical explanation of this sort? We here provide the most direct algorithm to date for answering this question in arbitrary prepare-and-measure experiments, and we provide open-access *Mathematica* code for answering it in practice. One need only give a finite set of quantum states and a finite set of quantum POVM elements as input, and the code determines if the statistics these generate by the Born rule can be explained classically—i.e., by a noncontextual ontological model for the operational scenario. It furthermore returns an explicit noncontextual model, if one exists. If there is no such model, the code determines an operational measure of nonclassicality, namely, the minimum amount of noise which would be required until a noncontextual model would become possible.

In the Supplemental Material [27], we generalize these ideas beyond quantum theory to the case of arbitrary generalized probabilistic theories (GPTs) [46,47] or fragments thereof, leveraging the fact that an operational scenario admits a noncontextual model if and only if the corresponding GPT admits a simplex embedding [4]. Indeed, the linear program we derive is simply a test of whether any valid simplex embedding (of any dimension) can be found, answering the challenge first posed in Ref. [4]. We furthermore prove an upper bound on the number of ontic states needed in any such classical explanation, namely, the square of the GPT dimension.

The Supplemental Material [27] also explains how our open-source code implements the linear program we develop herein.

A large number of previous works have studied the question of when a set of data admits a generalized noncontextual model [4–6,48–55]. Most closely related to our work are Refs. [5,49,50,52]. We elaborate on the relationships between these works in our conclusion and in Ref. [27].

For now, we simply note that the linear program (and dimension bound) that we derive here is closely related to an optimization problem introduced in Ref. [52]. However, Ref. [52] focuses on a proposed modification of generalized noncontextuality (which we criticize in the Supplemental Material [27]), and so the two approaches do not always return the same result.

Our Letter aims to be accessible and self-contained, in order to provide a tool for the quantum information and foundations communities to directly test for nonclassicality in their own research problems.

A linear program for deciding classicality.—We now set up the preliminaries required to state our linear program for testing whether the quantum statistics generated by given sets of quantum states and effects can be explained classically—i.e., by a noncontextual model for the operational scenario. The Supplemental Material [27] generalizes these ideas and results to arbitrary GPTs.

Consider any finite set of (possibly subnormalized [56]) quantum states Ω, and any finite set of quantum effects $\mathcal{E}$, living in the real vector space $\mathsf{Herm}[\mathcal{H}]$ of Hermitian operators on some finite dimensional Hilbert space $\mathcal{H}$. In general, neither the set of states nor the set of effects need span the full vector space $\mathsf{Herm}[\mathcal{H}]$, nor need the two sets span the same subspace of $\mathsf{Herm}[\mathcal{H}]$. Next, we introduce some useful mathematical objects related to Ω and $\mathcal{E}$.

Let us first focus on the case of states. We denote the subspace of $\mathsf{Herm}[\mathcal{H}]$ spanned by the states Ω by S_Ω. The inclusion map from S_Ω to $\mathsf{Herm}[\mathcal{H}]$ is denoted by I_Ω. In addition, we define the cone of positive operators that arises from Ω by

$$\mathsf{Cone}[\Omega] = \left\{\rho \middle| \rho = \sum_\alpha r_\alpha \rho_\alpha, \rho_\alpha \in \Omega, r_\alpha \in \mathbb{R}^+\right\} \subset S_\Omega. \quad (1)$$

This cone can also be characterized by its facet inequalities, indexed by $i = \{1, \ldots, n\}$, where n is necessarily finite as we start with a finite set of states (see, for example, McMullen's upper bound theorem [57]). These inequalities are specified by Hermitian operators $h_i^\Omega \in S_\Omega$ such that

$$\mathsf{tr}(h_i^\Omega v) \geq 0 \quad \forall\ i \Leftrightarrow v \in \mathsf{Cone}[\Omega]. \quad (2)$$

From these facet inequalities, one can define a linear map $H_\Omega\colon S_\Omega \to \mathbb{R}^n$, such that

$$H_\Omega(v) = (\mathsf{tr}(h_1^\Omega v), \ldots, \mathsf{tr}(h_n^\Omega v))^T \quad \forall\ v \in S_\Omega. \quad (3)$$

Note that the matrix elements of $H_\Omega(v)$ are all non-negative if and only if $v \in \mathsf{Cone}[\Omega]$. We denote entrywise non-negativity by $H_\Omega(v) \geq_e 0$ (to disambiguate from using ≥ 0 to represent positive semi-definiteness). Succinctly, we have

$$H_\Omega(v) \geq_e 0 \Leftrightarrow v \in \mathsf{Cone}[\Omega], \quad (4)$$

and so H_Ω is simply an equivalent characterization of the cone.

Consider now the set of effects $\mathcal{E}$. We denote the subspace of $\mathsf{Herm}[\mathcal{H}]$ spanned by $\mathcal{E}$ by $S_\mathcal{E}$, and the inclusion map from $S_\mathcal{E}$ to $\mathsf{Herm}[\mathcal{H}]$ by $I_\mathcal{E}$. In addition, we define the cone of positive operators that arises from $\mathcal{E}$ as

$$\mathsf{Cone}[\mathcal{E}] = \left\{\gamma \middle| \gamma = \sum_\beta r_\beta \gamma_\beta, \gamma_\beta \in \mathcal{E}, r_\beta \in \mathbb{R}^+\right\} \subset S_\mathcal{E}. \quad (5)$$

This cone can also be characterized by its facet inequalities, indexed by $j = \{1, \ldots, m\}$, where m is again finite, as we are considering a finite set of effects. These inequalities are specified by Hermitian operators $h_j^\mathcal{E}$ such that

$$\mathsf{tr}(h_j^\mathcal{E} w) \geq 0 \quad \forall\ j \Leftrightarrow w \in \mathsf{Cone}[\mathcal{E}]. \quad (6)$$

From these facet inequalities one can define a linear map $H_\mathcal{E}\colon S_\mathcal{E} \to \mathbb{R}^m$, such that

$$H_\mathcal{E}(w) = (\mathsf{tr}(w h_1^\mathcal{E}), \ldots, \mathsf{tr}(w h_m^\mathcal{E}))^T \quad \forall\ w \in S_\mathcal{E}. \quad (7)$$

This fully characterizes $\mathsf{Cone}[\mathcal{E}]$, since

$$H_\mathcal{E}(w) \geq_e 0 \Leftrightarrow w \in \mathsf{Cone}[\mathcal{E}]. \quad (8)$$

One can also pick an arbitrary orthonormal basis of Hermitian operators for each of the spaces $\mathsf{Herm}[\mathcal{H}]$, S_Ω, and $S_\mathcal{E}$, and represent I_Ω, $I_\mathcal{E}$, $H_\mathcal{E}$, and H_Ω as matrices with respect to these.

With these defined, we can now present the linear program which tests for classical explainability (i.e., simplex embeddability) of any set of quantum states and any set of quantum effects in terms of the matrices I_Ω, $I_\mathcal{E}$, H_Ω, and $H_\mathcal{E}$, defined above and computed from the set of states and set of effects.

Linear program 1.—The Born rule statistics obtained by composing any state-effect pair from Ω and $\mathcal{E}$ is classically explainable if and only if the following linear program is satisfiable:

$$\exists\, \sigma \geq_e 0, \qquad \text{an } m \times n \text{ matrix such that} \quad (9a)$$

$$I_\mathcal{E}^T \cdot I_\Omega = H_\mathcal{E}^T \cdot \sigma \cdot H_\Omega. \quad (9b)$$

Note that if Ω and $\mathcal{E}$ span the full vector space of Hermitian operators, then the linear program simplifies somewhat, as the l.h.s. of Eq. (9b) reduces to the identity map on $\mathsf{Herm}[\mathcal{H}]$. Note that satisfiability is only a function of the cones defined by Ω and by $\mathcal{E}$, and so no other features of the states and effects are relevant to their nonclassicality, as was also shown in Refs. [48,53]. A useful consequence of this fact is that Ω and $\mathcal{E}$ are classically explainable if and only if their convex hulls are also classically explainable.

Testing for the existence of such a σ is a linear program. In the repository [58], we give open-source *Mathematica* code for computing the relevant preliminaries and solving this linear program. The input to the code is simply a set of density matrices and a set of POVM elements (or, more generally, GPT state and effect vectors). In practice the code runs in a few seconds for values of n and m up to around 20.

In the case that a classical explanation does exist, the code will output a specification of an ontological model which represents the operational scenario in a noncontextual manner. This model can be computed from the matrix σ, as described in the Supplemental Material [27]. In particular, every density matrix in $\rho \in \Omega$ is represented in the ontological model by a probability distribution μ_ρ over some set of ontic states Λ, while every POVM element in $\varepsilon \in \mathcal{E}$ is represented by a response function ξ_ε—that is, a [0, 1]-valued function over Λ. Specifically, we compute a particular non-negative factorization $\sigma = \beta \cdot \alpha$, where $\alpha\colon \mathbb{R}^n \to \mathbb{R}^\Lambda \geq_e 0$ and $\beta\colon \mathbb{R}^\Lambda \to \mathbb{R}^m \geq_e 0$, and then construct linear maps $\tau_\Omega \coloneqq \alpha \cdot H_\Omega$ and $\tau_\mathcal{E} \coloneqq \beta^T \cdot H_\mathcal{E}$, and use these to define the epistemic states and response functions via

$$\mu_\rho(\lambda) \coloneqq [\tau_\Omega(\rho)]_\lambda \quad \text{and} \quad \xi_\varepsilon(\lambda) \coloneqq [\tau_\mathcal{E}(\varepsilon)]_\lambda \tag{10}$$

for all $\lambda \in \Lambda$. That these functions are all non-negative follows from the definition of H_Ω and $H_\mathcal{E}$ together with element-wise non-negativity of α and β; that they are suitably normalized follows from the manner in which the decomposition into α and β is chosen. In particular, the decomposition is constructed by taking $\beta = \sigma \cdot R$ and $\alpha = R^{-1}$, where R is a diagonal rescaling matrix which ensures that $\xi_\mathbb{1}(\lambda) = 1$ for all $\lambda \in \Lambda$ (see Supplemental Material [27], Sec. C.I for details). Note that other choices for the decomposition of $\sigma = \beta \cdot \alpha$ are possible, and that this nonuniqueness translates into a nonuniqueness of the ontological model.

In the case that no solution exists, one can ask how much depolarizing noise must be added to one's experiment until a solution becomes possible. This constitutes an operational measure of nonclassicality which we refer to as the *robustness of nonclassicality*. Finding the minimal amount r of noise is also a linear program:

Linear program 2.—Let r be the minimum depolarising noise that must be added in order for the statistics obtained by composing any state-effect pair from Ω and $\mathcal{E}$ to be classically explainable. It can be computed by the linear program:

minimize r such that

$$\exists\, \sigma \geq_e 0, \text{ an } m \times n \text{ matrix such that} \tag{11a}$$

$$r I_\mathcal{E}^T \cdot D \cdot I_\Omega + (1-r) I_\mathcal{E}^T \cdot I_\Omega = H_\mathcal{E}^T \cdot \sigma \cdot H_\Omega, \tag{11b}$$

where D is the completely depolarizing channel for the quantum system.

Again, the corresponding ontological model can be straightforwardly computed from the matrix σ found for the minimal value of r, and we give open-source code that returns both the value of r and the associated model.

We also discuss in the Supplemental Material [27] how one can easily adapt one's definition of robustness and the linear program for it to an arbitrary noise model.

Examples.—Here we present three examples of sets of states and effects, and we assess the classical-explainability of their statistics using our linear program. In the case where the statistics are not classical, we also compute the noise robustness. A fully detailed analysis of these examples (including the explicit calculation of the matrices H_Ω, $H_\mathcal{E}$, I_Ω, and $I_\mathcal{E}$), is given in the Supplemental Material [27]. These specific examples are chosen to illustrate particular features of our approach, as we discuss therein.

Example 1.—Consider the set of four quantum states

$$\Omega = \{|0\rangle\langle 0|, |1\rangle\langle 1|, |+\rangle\langle +|, |-\rangle\langle -|\} \tag{12}$$

on a qubit. In addition, consider the set of six effects

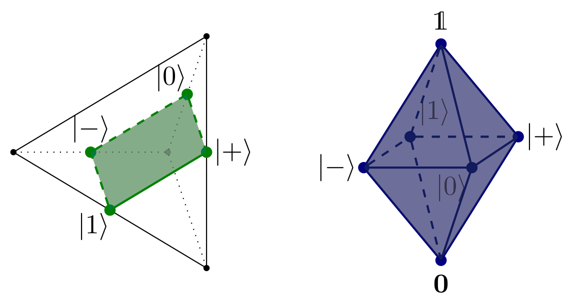
$$\mathcal{E} = \{|0\rangle\langle 0|, |1\rangle\langle 1|, |+\rangle\langle +|, |-\rangle\langle -|, \mathbb{1}_2, 0\}. \tag{13}$$

Next, consider the observable statistics—that is, the data that can be generated from any measurement constructed with these effects, when applied to any of these states.

Our linear program finds that these statistics admit a classical explanation. This is to be expected, as this scenario is a subtheory of the noncontextual toy theory of Ref. [59] (namely, that given by restricting to the real plane). Indeed, this is the model which our code returns, and is depicted in Fig. 1.

Example 2.—Consider the set of four quantum states

$$\Omega = \{|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|, |3\rangle\langle 3|\} \tag{14}$$



(a) Embedding of states (b) Embedding of effects

FIG. 1. Classical explanation for example 1. (a) Depiction of the states in Ω (green dots), embedded in a three-dimensional slice of a four-dimensional simplex. (b) Depiction of the effects in $\mathcal{E}$ (blue dots), embedded in a 3D slice of the 4D hypercube that is dual to the simplex in (a). Note that the convex hull of the effects happens to cover the entire hypercube in this particular slice. The simplex in (a) can be viewed as the set of probability distributions over a 4-element set Λ of ontic states (black dots), while the hypercube in (b) can be viewed as the set of logically possible response functions for Λ. Hence, this simplex embedding corresponds to a noncontextual ontological model for—and hence [4] a classical explanation of—the operational scenario.

on a four-dimensional quantum system. In addition, consider the set of six effects

$$\mathcal{E} = \{|0\rangle\langle 0| + |1\rangle\langle 1|, |1\rangle\langle 1| + |2\rangle\langle 2|, |2\rangle\langle 2| + |3\rangle\langle 3|, |3\rangle\langle 3| + |0\rangle\langle 0|, \mathbb{1}_4, 0\}. \tag{15}$$

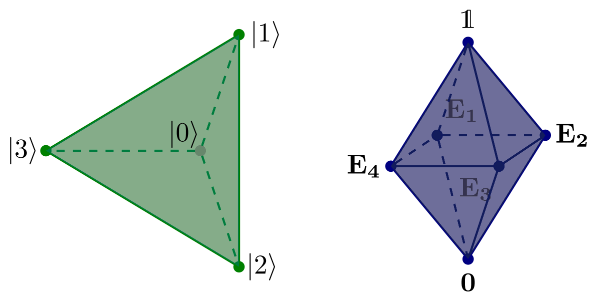
Notably, the states and effects in this example do not span the same vector space. Still, our linear program also finds that the statistical data that arise from this admits a classical explanation. This is a useful sanity check, since all the states and effects are diagonal in the same basis. We provide a depiction of the classical model which our code returns for this scenario in Fig. 2.

Example 3.—Our third example is obtained from the first example by rotating all of the effects by an angle of $(\pi/4)$ about the σ_y axis. (This is the set of states and effects relevant for parity-oblivious multiplexing [15].) In this case, our linear program finds that there is no classical explanation of the observable statistics. Moreover, it finds that the depolarizing-noise robustness for these states and effects is $r = 1 - (1/\sqrt{2}) \sim 0.3$. In Fig. 3 we depict the classical model for the case of depolarization at this noise threshold.

Related linear programs.—We reiterate that the core of our linear program is closely related to the linear program introduced in Sec. 4.2 of Ref. [52] as specialized to the polytopic case (that we consider here) in Sec. 4.3 of Ref. [52]. However, the approach of Ref. [52] differs from ours in a critical preprocessing step, and so its assessment of classicality differs from ours in some examples. Indeed, their proposal deems example 2 nonclassical, while our approach deems it classical. But, the "nonclassical" verdict is clearly mistaken, since all the states and effects in that example are simultaneously diagonalizable. Still, we emphasize that the mathematical tools of Ref. [52] are quite useful and applicable to our notion of classicality, and indeed even extend some results to non-polytopic GPTs (although in this case, testing for nonclassicality is likely not a linear program) via inner and outer polytopic approximations as discussed in Sec. 4.4 of Ref. [52].

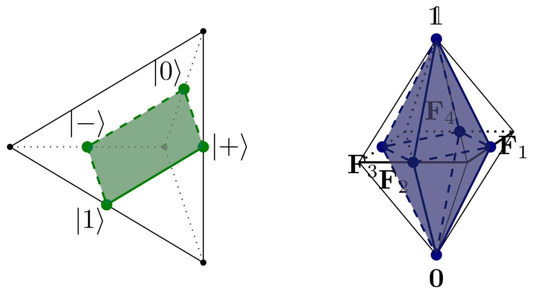
Reference [50] also presented a linear programming approach which could determine if a prepare-measure scenario admits a noncontextual model or not. In that work, however, the input to the linear program required the specification of a set of operational equivalences for the states and another set for the effects. In contrast, in the current work, the input to the algorithm is simply a set of quantum (or GPT) states and effects. The full set of operational equivalences that hold among states and among effects are derivable from this input; however, one need not consider them explicitly. The linear program we present here determines if there is a noncontextual model with respect to *all* of the operational equivalences that hold in quantum theory (or within the given GPT).

Reference [49] provided another linear programming approach to testing noncontextuality in the context of a particular class of prepare-measure scenarios; namely, those wherein all operational equivalences arise from different ensembles of preparation procedures, all of which define the same average state. Using the flag-convexification technique of Refs. [48,53], we suspect that *all* prepare-measure



(a) Embedding of states (b) Embedding of effects

FIG. 2. Classical explanation for example 2. (a) Depiction of the states in Ω (green dots), embedded in a 3D slice of a 4D simplex. (b) Depiction of the effects in $\mathcal{E}$ (blue dots), embedded in a 3D slice of the 4D hypercube that is dual to the simplex in (a). Note that the convex hull of the states (effects) happens to cover the entire simplex (hypercube) in this particular slice. Exactly as in the last example, this simplex embedding corresponds to a noncontextual ontological model for—and hence [4] a classical explanation of—the operational scenario.



(a) Embedding of states (b) Embedding of effects

FIG. 3. Classical explanation for example 3, when depolarized by $r = 1 - (1/\sqrt{2})$. (a) Depiction of the states in Ω (green dots), embedded in a 2D slice of a 4D simplex. (b) Depiction of the effects in $\mathcal{E}$ (blue dots), embedded in a 3D slice of the 4D hypercube that is dual to the simplex in (a). Exactly as in the last example, this simplex embedding corresponds to a noncontextual ontological model for—and hence [4] a classical explanation of—the depolarized operational scenario. If the depolarization was less strong, then such a noncontextual ontological model would not exist. Visually, we can get some intuition for this by observing that if we grow either the green square or the blue octahedron, then we would end up with the states or effects lying outside of the simplex or hypercube.

scenarios can be transformed into prepare-measure scenarios of this particular type, in which case the linear program from Ref. [49] would be as general as the approach we have discussed herein. However, this remains to be proven.

An interesting open question is to determine the relative efficiency of these algorithms.

Closing remarks.—Our arguments in the Supplemental Material [27] demonstrate that if a noncontextual model exists for a scenario, then there also exists a model with $\mathsf{dim}[S_\Omega]\,\mathsf{dim}[S_\mathcal{E}] \leq \mathsf{dim}[\mathcal{H}]^2$ ontic states (or less), This bound was first proven in Ref. [52] by similar arguments. It is not yet clear if this bound is tight.

Additionally, our arguments in the Supplemental Material [27] hinge on the existence of a particular kind of decomposition of the identity channel. The arguments proving a structure theorem for noncontextual models in Ref. [6] hinged on a similar decomposition of the identity channel, and it would be interesting to investigate this connection further. We hope that a synthesis of the algorithmic techniques herein with the compositional techniques of Refs. [6,60] might lead to algorithms for testing nonclassicality in prepare-transform-measure scenarios and eventually in arbitrary circuits.

In Ref. [54], the definition of simplex embedding was generalized to embeddings into arbitrary GPTs. It would be interesting to investigate whether similar programs (albeit most likely not linear ones [61].) could be developed for testing for such embeddings.

Finally, we note that our linear program and open source implementation are ideally suited for proving nonclassicality in real experiments [62], especially when coupled with theory-agnostic tomography techniques [63,64].

We acknowledge useful feedback from Rob Spekkens. J. H. S. was supported by the National Science Centre, Poland (Opus project, Categorical Foundations of the Non-Classicality of Nature, Project No. 2021/41/B/ST2/03149). A. B. S., D. S., and V. P. R. were supported by the Foundation for Polish Science (IRAP project, ICTQT, Contract No. MAB/2018/5, co-financed by EU within Smart Growth Operational Programme). E. W. was supported by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported in part by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Colleges and Universities. All diagrams were made using TikZiT.

*john.h.selby@gmail.com
†ewolfe@perimeterinstitute.ca
‡davidschmid10@gmail.com
§ana.sainz@ug.edu.pl
‖prettirossi.vinicius@gmail.com

[1] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, Phys. Rev. A **71**, 052108 (2005).
[2] S. Kochen and E. Specker, The problem of hidden variables in quantum mechanics, J. Math. Mech. **17**, 59 (1967), also available from the Indiana University Mathematics Journal
[3] R. W. Spekkens, The ontological identity of empirical indiscernibles: Leibniz's methodological principle and its significance in the work of Einstein, arXiv:1909.04628.
[4] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, Characterization of noncontextuality in the framework of generalized probabilistic theories, PRX Quantum **2**, 010331 (2021).
[5] F. Shahandeh, Contextuality of general probabilistic theories, PRX Quantum **2**, 010330 (2021).
[6] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, A structure theorem for generalized-noncontextual ontological models, arXiv:2005.07161.
[7] R. W. Spekkens, Negativity and contextuality are equivalent notions of nonclassicality, Phys. Rev. Lett. **101**, 020401 (2008).
[8] R. D. Baldijão, R. Wagner, C. Duarte, B. Amaral, and M. T. Cunha, Emergence of noncontextuality under quantum Darwinism, PRX Quantum **2**, 030351 (2021).
[9] R. Kunjwal and R. W. Spekkens, From the Kochen-Specker theorem to noncontextuality inequalities without assuming determinism, Phys. Rev. Lett. **115**, 110403 (2015).
[10] I. Marvian, Inaccessible information in probabilistic models of quantum systems, non-contextuality inequalities and noise thresholds for contextuality, arXiv:2003.05984.
[11] J. S. Bell, On the Einstein Podolsky Rosen paradox, Physics **1**, 195 (1964).
[12] R. Kunjwal and R. W. Spekkens, From statistical proofs of the Kochen-Specker theorem to noise-robust noncontextuality inequalities, Phys. Rev. A **97**, 052110 (2018).
[13] M. F. Pusey, Anomalous weak values are proofs of contextuality, Phys. Rev. Lett. **113**, 200401 (2014).
[14] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts, Phys. Rev. A **100**, 042116 (2019).
[15] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Preparation contextuality powers parity-oblivious multiplexing, Phys. Rev. Lett. **102**, 010401 (2009).
[16] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, Optimal bounds for parity-oblivious random access codes, New J. Phys. **18**, 045003 (2016).
[17] A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai, Parity oblivious d-level random access codes and class of noncontextuality inequalities, Quantum Inf. Process. **18**, 111 (2019).
[18] D. Saha, P. Horodecki, and M. Pawłowski, State independent contextuality advances one-way communication, New J. Phys. **21**, 093057 (2019).
[19] S. A. Yadavalli and R. Kunjwal, Contextuality in entanglement-assisted one-shot classical communication, Quantum **6**, 839 (2022).
[20] D. Schmid, H. Du, J. H. Selby, and M. F. Pusey, Uniqueness of noncontextual models for stabilizer subtheories, Phys. Rev. Lett. **129**, 120403 (2022).

[21] D. Schmid and R. W. Spekkens, Contextual advantage for state discrimination, Phys. Rev. X **8**, 011015 (2018).

[22] K. Flatt, H. Lee, C. R. I. Carceller, J. B. Brask, and J. Bae, Contextual advantages and certification for maximum-confidence discrimination, PRX Quantum **3**, 030337 (2022).

[23] S. Mukherjee, S. Naonit, and A. K. Pan, Discriminating three mirror-symmetric states with a restricted contextual advantage, Phys. Rev. A **106**, 012216 (2022).

[24] J. Shin, D. Ha, and Y. Kwon, Quantum contextual advantage depending on nonzero prior probabilities in state discrimination of mixed qubit states, Entropy **23**, 1583 (2021).

[25] M. Lostaglio and G. Senno, Contextual advantage for state-dependent cloning, Quantum **4**, 258 (2020).

[26] M. Lostaglio, Certifying quantum signatures in thermodynamics and metrology via contextuality of quantum linear response, Phys. Rev. Lett. **125**, 230603 (2020).

[27] See the Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.132.050202 for the Appendices referred to along this Letter, which include Refs. [28–45].

[28] F. Shahandeh, M. J. W. Hall, and T. C. Ralph, Measurement-device-independent approach to entanglement measures, Phys. Rev. Lett. **118**, 150505 (2017).

[29] M. F. Pusey, L. Del Rio, and B. Meyer, Contextuality without access to a tomographically complete set, arXiv:1904.08699.

[30] R. W. Spekkens, Quasi-quantization: Classical statistical theories with an epistemic restriction, in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer Netherlands, Dordrecht, 2016), pp. 83–135.

[31] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Reconstruction of Gaussian quantum mechanics from Liouville mechanics with an epistemic restriction, Phys. Rev. A **86**, 012103 (2012).

[32] L. Catani, M. Leifer, D. Schmid, and R. W. Spekkens, Why interference phenomena do not capture the essence of quantum theory, Quantum **7**, 1119 (2023).

[33] M. Müller, Probabilistic theories and reconstructions of quantum theory, SciPost Phys. Lect. Notes 028 (2021).

[34] L. Lami, Non-classical correlations in quantum mechanics and beyond, arXiv:1803.02902.

[35] M. Plávala, General probabilistic theories: An introduction, Phys. Rep. **1033**, 1 (2023).

[36] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Quantum from principles, in *Fundamental Theories of Physics* (Springer Netherlands, Dordrecht, 2015), pp. 171–221.

[37] L. Hardy, Reformulating and reconstructing quantum theory, arXiv:1104.2066.

[38] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Probabilistic theories with purification, Phys. Rev. A **81**, 062348 (2010).

[39] B. Coecke, T. Fritz, and R. W. Spekkens, A mathematical theory of resources, Inf. Comput. **250**, 59 (2016).

[40] T. Gonda and R. W. Spekkens, Monotones in general resource theories, Compositionality **5**, 7 (2023).

[41] V. P. Rossi, D. Schmid, J. H. Selby, and A. B. Sainz, Contextuality with vanishing coherence and maximal robustness to dephasing, Phys. Rev. A **108**, 032213 (2023).

[42] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Nonlocal correlations as an information-theoretic resource, Phys. Rev. A **71**, 022101 (2005).

[43] G. Kimura, The Bloch vector for n-level systems, Phys. Lett. A **314**, 339 (2003).

[44] T. S. Motzkin, H. Raiffa, G. L. Thompson, and R. M. Thrall, The double description method, Contrib. Theory Games **2**, 51 (1953).

[45] D. Avis and K. Fukuda, A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra, in *Proceedings of the Seventh Annual Symposium on Computational Geometry* (1991), pp. 98–104.

[46] L. Hardy, Quantum theory from five reasonable axioms, arXiv:quant-ph/0101012.

[47] J. Barrett, Information processing in generalized probabilistic theories, Phys. Rev. A **75**, 032304 (2007).

[48] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing nonclassicality, Phys. Rev. A **107**, 062203 (2023).

[49] A. Krishna, R. W. Spekkens, and E. Wolfe, Deriving robust noncontextuality inequalities from algebraic proofs of the Kochen–Specker theorem: The Peres-Mermin square, New J. Phys. **19**, 123031 (2017).

[50] D. Schmid, R. W. Spekkens, and E. Wolfe, All the noncontextuality inequalities for arbitrary prepare-and-measure experiments with respect to any fixed set of operational equivalences, Phys. Rev. A **97**, 062103 (2018).

[51] M. F. Pusey, Robust preparation noncontextuality inequalities in the simplest scenario, Phys. Rev. A **98**, 022112 (2018).

[52] V. Gitton and M. P. Woods, Solvable criterion for the contextuality of any prepare-and-measure scenario, Quantum **6**, 732 (2022).

[53] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Contextuality without incompatibility, Phys. Rev. Lett. **130**, 230201 (2023).

[54] M. P. Müller and A. J. P. Garner, Testing quantum theory by generalizing noncontextuality, Phys. Rev. X **13**, 041001 (2023).

[55] A. Chaturvedi, M. Farkas, and V. J. Wright, Characterising and bounding the set of quantum behaviours in contextuality scenarios, Quantum **5**, 484 (2021).

[56] This allows one to describe states which are not prepared deterministically, as happens, e.g., when the preparation is part of a probabilistic source or is the result of remote steering.

[57] P. McMullen, The maximum numbers of faces of a convex polytope, Mathematika **17**, 179 (1970).

[58] Open-source Mathematica code for implementing the linear program introduced in this Letter (2022), https://github.com/eliewolfe/SimplexEmbedding. See also https://github.com/pjcavalcanti/SimplexEmbeddingGPT for a PYTHON implementation of the program.

[59] R. W. Spekkens, Evidence for the epistemic view of quantum states: A toy theory, Phys. Rev. A **75**, 032110 (2007).

[60] D. Schmid, J. H. Selby, and R. W. Spekkens, Unscrambling the omelette of causation and inference: The framework of causal-inferential theories, arXiv:2009.03297.

[61] E.g., in the quantum case these are typically SDP hierarchies [55].
[62] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, An experimental test of noncontextuality without unphysical idealizations, Nat. Commun. **7**, 1 (2016).
[63] M. J. Grabowecky, C. A. J. Pollack, A. R. Cameron, R. W. Spekkens, and K. J. Resch, Experimentally bounding deviations from quantum theory for a photonic three-level system using theory-agnostic tomography, Phys. Rev. A **105**, 032204 (2022).
[64] M. D. Mazurek, M. F. Pusey, K. J. Resch, and R. W. Spekkens, Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories, PRX Quantum **2**, 020302 (2021).

Supplemental material to "A linear program for testing nonclassicality and an open-source implementation"

John H. Selby,[1, *] Elie Wolfe,[2, †] David Schmid,[1, ‡] Ana Belén Sainz,[1, §] and Vinicius P. Rossi[1, ¶]

[1] *International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-309 Gdańsk, Poland*
[2] *Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario Canada N2L 2Y5*

(Dated: January 4, 2024)

CONTENTS

A. Related work 1
I. Shahandeh's proposed notion of classicality 1
II. Gitton-Woods's proposed notion of classicality 1
III. Müller-Garner's notion of simulability 2

B. Formal definitions 2
I. Generalized probabilistic theories 2
II. GPT fragments 4
III. Accessible GPT fragments 5
IV. Classical explainability of accessible GPT fragments 6

C. Derivation of the linear program 7
I. From simplicial-cone embeddings to simplex embeddings 9
II. From simplex embeddings to ontological models 10

D. An operational measure of nonclassicality 10

E. Bounding the number of ontic states 11

F. Worked Examples 12
I. Example 1 12
II. Example 2 14
III. Example 3 14
IV. Example 4 15

G. Open-source code 16
I. How to use the code 16
II. Useful matrix formulations 17
III. Internals of the code 17

References 18

A. RELATED WORK

The examples we examined in Section F are useful for comparing generalized noncontextuality with two other purported notions of classicality, introduced by Shahandeh in Ref. [1] and by Gitton and Woods in Ref. [2]. After this, we also comment on a third related idea from Ref. [3].

I. Shahandeh's proposed notion of classicality

In Ref. [4], Shahandeh proposes that an operational theory whose GPT has dimension d should be deemed classically-explainable if and only if it admits of a noncontextual model *whose ontic state space contains exactly d ontic states*. Clearly, this implies generalized noncontextuality, but it is a strictly more stringent condition. This can be seen from the fact that Example 1 from the main text admits of a generalized noncontextual model, and noting that this model *necessarily* requires 4 ontic states [5], while $d = 3$. Our technique deems this example classically-explainable (and provides a noncontextual model for it), while the example is considered not classically-explainable by Shahandeh's definition.

Despite Shahandeh's arguments in Ref. [4] for including this dimensional restriction (as well as our own attempts to find a motivation for it, e.g., by appealing to some variant of Leibniz's principle), we have not yet seen any compelling motivations for it.

II. Gitton-Woods's proposed notion of classicality

In Ref. [2], Gitton and Woods propose a novel notion of nonclassicality. To understand their proposal, first recall that generalized noncontextuality only implies constraints on the ontological representations of operational processes which are genuinely operationally equivalent—that is, those that give exactly the same predictions for any operational scenario in which they appear. For example, two preparation procedures are only operationally equivalent if they give the same statistics for the outcomes *of all possible measurements*, and it is only in this case that one has reason to expect that their ontological representations be identical as stochastic processes (as the principle of noncontextuality demands).

In contrast, Ref. [2] proposes that we drop this requirement, and instead demand that operational processes which make the same predictions *for the particular measurements performed in a given experiment* should *also* have the same ontological representations as stochastic processes, even if these operational processes

* john.h.selby@gmail.com
† ewolfe@perimeterinstitute.ca
‡ davidschmid10@gmail.com
§ ana.sainz@ug.edu.pl
¶ prettirossi.vinicius@gmail.com

are ultimately distinguishable by other procedures (not considered in the given experiment). As an example, this approach would demand that the ontological representations of $|0\rangle + i|1\rangle$ and $|0\rangle - i|1\rangle$ must be identical in any experiment wherein the only measurements performed on the system were confined to the $X-Z$ plane, since all such measurements give the same statistics on these two states.

The notion of classicality that results in Ref. [2] is hence distinct from generalized noncontextuality. One can see this explicitly from a particular example: it deems our Example 2 nonclassical[1], even though there exists a generalized noncontextual ontological model for it, as we demonstrated in Section F.

Recall, however, that our Example 2 involved only states and effects which are diagonal in a fixed basis. Given that simultaneously diagonalizability is widely considered to be a strong notion of classicality, it is clear that Gitton-Woods' proposal is problematic.

In addition, their approach does not respect the well motivated constraint that any subset of states and effects from a classically-explainable theory must also be classically-explainable. (One can verify this by considering our Example 2 as a sub-fragment of the full set of quantum states and effects in the computational basis of a 4-dimensional Hilbert space, and noting that Gitton-Woods' proposal deems this strictly larger scenario classical[2]).

These examples illustrate the fact that one's assessment of nonclassicality, at least in the sense of generalised noncontextuality, is only as good as one's knowledge of the true operational equivalences. If one quotients with respect to a set of operational procedures which is *not* tomographically complete for the system under consideration, then one can mistakenly conclude that one's experiment does not admit of a classical explanation. For more on this, see the discussions of tomographic completeness in Refs. [6**?** , 7], and see also Refs. [8, 9] for methods for obtaining evidence that one truly has tomographically complete ways of probing a given system.

The mathematical tools developed by Gitton and Woods can be reconsidered as *sufficient* (but not necessary) conditions for classical-explainability (defined by existence of a generalized noncontextual model). Moreover, other elements of the analysis in Ref. [2] are quite useful; as we noted above, for example, Ref. [2] essentially derives the same linear program that we developed (independently) and presented here and provides useful techniques for polytopic approximations to situations involving infinite sets of states and effects.

[1] To see that it is deemed nonclassical by the approach of Ref. [2], it suffices to note that if one quotients its (three-dimensional) realized state space with respect to its (two-dimensional) realized effect space, the resulting GPT is Boxworld, as in our fourth example, which we show is not simplex-embeddable.

[2] This can be seen from the fact that their approach coincides with ours when the states and effects span the same space.

III. Müller-Garner's notion of simulability

Finally, we comment on some related ideas from Ref. [3], in which Müller and Garner define a notion of 'classical simulation'. A simulation associates to each state (effect) of a given GPT a *set* of states (effects) in a simplicial GPT, such that a few natural conditions hold (e.g., probabilities are reproduced, and convex mixture is respected in the appropriate sense).

Müller and Garner do not claim that this is a notion of classicality; indeed, they reprove a result by Holevo establishing that every GPT has a classical 'simulation' of this sort. Hence, this notion of simulability cannot form the basis of a useful notion of classicality, as it does not establish a meaningful dividing line between classical and nonclassical phenomena.

As an aside, it is interesting to note that the mathematical object which simulates a given GPT can be viewed as an epistemically restricted ontological theory [10–13]. This is analogous to how any simplex-embeddable GPT can be viewed as an epistemically restricted ontological theory. However, in the latter case, the epistemic states and response functions allowed by the epistemic restriction necessarily span the same vector space[3], whereas in the former, this need not be the case.

Ontological theories which are epistemically restricted in this manner (where the epistemic states and response functions do not span the same vector space) are consequently able to reproduce *all* operational theories, including those which do not admit of any noncontextual representation. An interesting question for future work would be to provide independent physical motivations for rejecting epistemic restrictions of this sort as natural classical explanations.

B. FORMAL DEFINITIONS

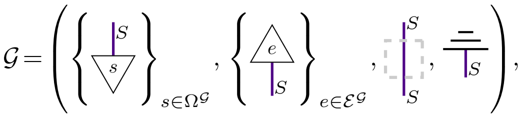
I. Generalized probabilistic theories

In the main text, we presented our results using the formalism of quantum theory. In the following, we state and prove our results in a more general manner which does not assume the validity of quantum theory. We do this within the framework of generalized probabilistic theories (GPTs).

Ref. [14] contains a comprehensive introduction to the particular formalization of GPTs which we use here. In addition, see Refs. [15–17] for reviews of GPTs more broadly, and Refs. [18, 19] for the diagrammatic formalism for GPTs that we use here. In brief, a GPT describes a possible theory of the world, as characterized by its operational statistics [20, 21]. By ranging over different GPTs, then, one ranges over a landscape of possible ways the world might be.

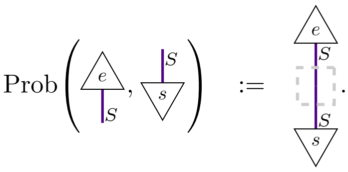
[3] One can verify that this property is satisfied in all of the epistemically restricted theories studied to date [10, 11], as required by the fact that they constitute noncontextual models.

We now briefly review the GPT description of prepare-measure scenarios, which are the focus of our manuscript. In this context, a GPT is formally defined by a quadruple


(1)

where $\Omega^{\mathcal{G}}$ is a convex set of GPT states which span a real vector space S, and $\mathcal{E}^{\mathcal{G}}$ is a convex set of GPT effects which span the dual space S^*. Each GPT state represents an operational preparation procedure—possibly one that occurs with non-unit probability[4], and each GPT effect represents an operational measurement procedure together with the observation of a particular outcome. We follow the standard convention of assuming that the sets $\Omega^{\mathcal{G}}$ and $\mathcal{E}^{\mathcal{G}}$ are finite-dimensional, convex, and compact. The quadruple also specifies a probability rule (via the identity map $\mathbb{1}_S$—see Eq. (2)), and a unit effect $\bar{\top}_S \in \mathcal{E}^{\mathcal{G}}$. These are in fact both redundant in the case of standard GPTs, but we have included them here to highlight the fact that standard GPTs are special cases of accessible GPT fragments, a concept we introduced in Ref. [14] and which we review in the next section.

A measurement in a GPT is a set of effects (one for each possible outcome) summing to the privileged unit effect, $\bar{\top}_S$. In any measurement containing a GPT effect $e \in \mathcal{E}^{\mathcal{G}}$, the probability of the outcome corresponding to that effect arising, given a preparation of the system in a state described by the GPT state $s \in \Omega^{\mathcal{G}}$, is given by


(2)

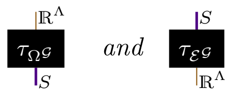
The set of states and effects in any valid GPT must satisfy a number of constraints, three of which we highlight here:

1. The principle of *tomography* [20, 21] must be satisfied. This means that all states and effects can be uniquely identified by the predictions they generate. Formally, for GPT states it means that $e(s_1) = e(s_2)$ for all $e \in \mathcal{E}^{\mathcal{G}}$ if and only if $s_1 = s_2$; for GPT effects, it means that $e_1(s) = e_2(s)$ for all $s \in \Omega^{\mathcal{G}}$ if and only if $e_1 = e_2$.
2. For every state $s \in \Omega^{\mathcal{G}}$, it holds that $\frac{1}{\bar{\top}_S(s)} s \in \Omega^{\mathcal{G}}$. That is, for every state in the GPT, the normalised counterpart is also in the GPT. This is an important constraint first highlighted in Ref. [22].
3. For all $e \in \mathcal{E}^{\mathcal{G}}$, there exists $e^\perp \in \mathcal{E}^{\mathcal{G}}$ such that $e + e^\perp = \bar{\top}_S$.

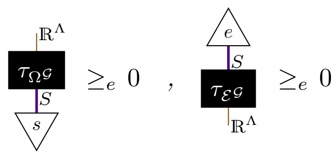
We highlight these in particular as they are relevant for this manuscript. Notably, we will relax the first two conditions in the following section.

In this paper we are interested in whether or not a given GPT, or an experiment performed within a given GPT, is classically explainable. In Refs. [5, 14], it was shown that the appropriate notion of classical-explainability is the notion of simplex-embeddability. This geometric criterion deems a GPT *classically-explainable* if its state space can be embedded into a simplex (of any dimension) and its effect space can be embedded into the dual to the simplex, such that probabilities are preserved. This notion is motivated by the fact that the existence of a simplex embedding for a GPT is equivalent to the existence of a generalized noncontextual ontological model for any operational scenario which leads (through quotienting by operational equivalences) to the GPT. In turn, recall (e.g., from our introduction) that there are many motivations for taking generalized noncontextuality as one's notion of classical explainability for operational scenarios. Simplex-embeddability can also be motivated as a notion of classical-explainability by the independent consideration that simplicial GPTs are the standard way of capturing strictly classical theories—i.e., those wherein all possible measurements are compatible. We refer the reader to Refs. [5, 14] for more details on this and on the closely related notion of simplicial-cone embedding.

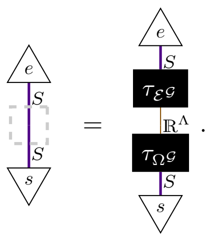
Definition 1 (Simplicial-cone embeddings and simplex embeddings of a GPT)**.** *A* simplicial-cone embedding, $\tau_{\mathcal{G}}$, *of a GPT, $\mathcal{G}$, is defined by a set of ontic states Λ and a pair of linear maps*


(3)

such that for all $s \in \Omega^{\mathcal{G}}$ and for all $e \in \mathcal{E}^{\mathcal{G}}$ we have


(4)

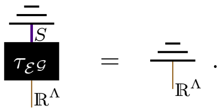
and such that


(5)

A simplicial-cone embedding is said to be a simplex

[4] That is, we take $\Omega^{\mathcal{G}}$ to include subnormalised states, representing preparation procedures which fail with some nonzero probability, as this will be convenient later in the paper.

embedding *if it moreover satisfies*



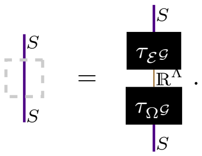
. (6)

Although a simplex embedding must satisfy this additional constraint, we proved in Ref. [14] that a simplicial-cone embedding exists if and only if a simplex embedding exists. We expand on this in Section C I.

We note also that a simplex-embedding of a GPT is equivalent to an ontological model of a GPT [5, 23]. It follows, then, that an operational theory (or scenario or experiment) admits of a noncontextual ontological model if and only if the associated GPT (or GPT fragment, see next subsection) admits of an ontological model.

Here and throughout, we will denote physical processes in white, and mathematical processes (like embedding, projection, and inclusion maps) in black.

Note that since states are spanning for S and effects are spanning for S^*, we can equivalently write Eq. (5), which expresses the constraint that the operational data is reproduced by the embedding, as



. (7)

II. GPT fragments

In a standard GPT (as defined in the previous section), *all* states and effects which are taken to be physically possible given one's theory of the world are required to be included in the sets $\Omega^{\mathcal{G}}$ and $\mathcal{E}^{\mathcal{G}}$. When applying the framework of GPTs to describe *particular experiments* rather than possible theories of the world, however, one must drop this requirement.

From a practical perspective, a specific prepare-measure experiment can be described simply as a subset $\Omega^{\mathcal{F}} \subseteq \Omega^{\mathcal{G}}$ representing the preparation procedures in the experiment and a subset $\mathcal{E}^{\mathcal{F}} \subseteq \mathcal{E}^{\mathcal{G}}$ of effects representing the measurement outcomes in the experiment. We will refer to this object as a *GPT fragment*. More formally:

Definition 2 (GPT fragment)**.** *A GPT fragment, $\mathcal{F}$, is specified by the underlying GPT, $\mathcal{G}$, together with a designated subset of states $\Omega^{\mathcal{F}} \subseteq \Omega^{\mathcal{G}}$ and of effects $\mathcal{E}^{\mathcal{F}} \subseteq \mathcal{E}^{\mathcal{G}}$.*

Critically, the sets of states and effects in a GPT fragment ($\Omega^{\mathcal{F}}$ and $\mathcal{E}^{\mathcal{F}}$) need not satisfy all the constraints that a GPT must satisfy. In particular, (i) the set of state vectors and effect covectors in a GPT fragment need not be tomographically complete for each other (i.e., they need not span the same vector space and its dual, respectively), and (ii) the set of state vectors in a GPT fragment may contain subnormalized states whose normalised counterparts are not in the GPT fragment.

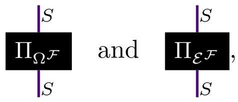
One can then ask whether the fragment (as opposed to the underlying GPT) is classically-explainable. The appropriate notion of classical-explainability for GPT fragments follows immediately from the notion for the underlying GPT, namely Definition 1:

Definition 3 (Simplicial-cone embeddings and simplex embeddings of GPT fragments)**.** *Definition 1, but where one replaces $\Omega^{\mathcal{G}}$ with $\Omega^{\mathcal{F}}$ and replaces $\mathcal{E}^{\mathcal{G}}$ with $\mathcal{E}^{\mathcal{F}}$.*

Note that any fragment $\mathcal{F}$ of a classically-explainable underlying GPT $\mathcal{G}$ is necessarily also classically-explainable; contrapositively, if a fragment is not classically-explainable, then neither is the underlying GPT.

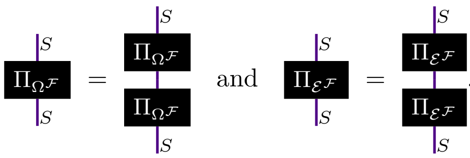
Note that $\Omega^{\mathcal{F}}$ and $\mathcal{E}^{\mathcal{F}}$ are not necessarily spanning for the underlying GPT vector space S and dual space S^*, respectively. As such, one can no longer derive Eq. (7) as an equivalent way to capture the constraint that the operational predictions be reproduced (as in Eq. (5)). However, we can derive an analogous condition by introducing some projection maps. Although this is not necessary at this stage, these maps will be useful in the next section.

Define a particular pair of idempotent linear maps



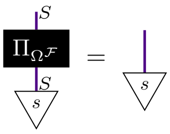
, (8)

where idempotence means that



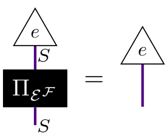
. (9)

The defining feature of these idempotents is that they characterize the subspaces of states and effects in the fragment via



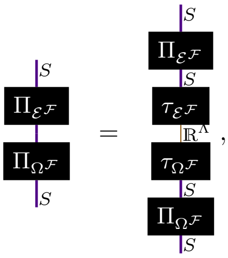
$\iff \quad s \in \mathsf{Span}[\Omega^{\mathcal{F}}]$ (10)

and



$\iff \quad e \in \mathsf{Span}[\mathcal{E}^{\mathcal{F}}].$ (11)

Although more than one idempotent map may satisfy these constraints, they are related by reversible linear maps relating two different choices of bases for S, and so we will see that the results hold for any choice satisfying these conditions. Then, Eq. (7) can be more generally expressed as


(12)

which now directly applies to GPT fragments as well. These idempotents and this equivalent characterisation of Eq. (5) will be useful in the following section.

III. Accessible GPT fragments

As one can see from Definition 2, a GPT fragment is explicitly defined *with respect* to an underlying GPT.

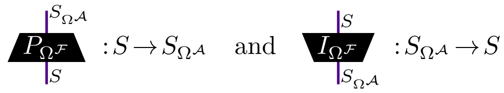
From a theorist's perspective, however, it can be convenient to work with an 'intrinsic' characterisation of an experiment, rather than viewing it as a fragment living inside an underlying GPT. That is, it is often useful to view one's subsets of states and effects as living in the vector spaces which they span, rather than the vector space of the underlying GPT (a vector space that will generally be of larger dimension). This resulting object has been termed an *accessible GPT fragment* [14].

The definition of an accessible GPT fragment in Ref. [14] also incorporates a closure of the state space and of the effect space under classical processings—convex mixtures, coarse-grainings of outcomes, and so on.[5] Thus, they represent all states and effects that are *accessible* given the laboratory devices in question (but, like with GPT fragments, they need not represent all states and effects that are physically possible in the underlying theory). As a consequence of this closure under classical processings, all accessible GPT fragments share some geometric structure; however, this additional structure is not needed for proving our results, so we simply refer the reader to Ref. [14] for more discussion of this.

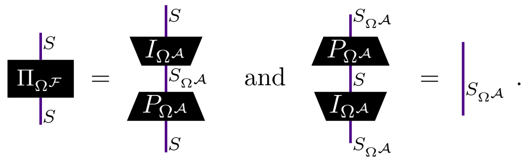
In summary, accessible GPT fragments are simply GPT fragments, but represented in their native vector spaces, and closed under classical processing. As with GPT fragments, the set of state vectors and effect covectors in an accessible GPT fragment need not be tomographically complete for each other (i.e., they need not span the same vector space and its dual), and the set of state vectors may contain subnormalized states whose normalised counterparts are not in the accessible GPT fragment.

[5] Note that the choice to incorporate this closure is in some sense optional; one could also study objects like accessible GPT fragments, but without this closure.

In order to formalize this move from the GPT fragments of the previous section to accessible GPT fragments, we will make use of the idempotent maps that we introduced. In particular, let us define a "splitting" of these idempotents as follows. In the case of $\Pi_{\Omega^\mathcal{F}}$, this means finding a vector space $S_{\Omega^\mathcal{A}}$ and a pair of linear maps


(13)

such that

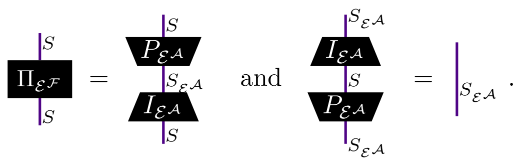

(14)

In particular, these conditions mean that $S_{\Omega^\mathcal{A}} \cong \mathsf{Span}[\Omega^\mathcal{F}]$, so we will think of $S_{\Omega^\mathcal{A}}$ as the vector space of states for the accessible GPT fragment. The map $P_{\Omega^\mathcal{A}}$ is then a projector mapping states *viewed as vectors within the underlying GPT* to states *viewed as vectors in the accessible GPT fragment*. Meanwhile, the map $I_{\Omega^\mathcal{A}}$ is an inclusion map taking states *viewed as vectors within the accessible GPT fragment* to states *viewed as vectors within the GPT*. Note that splitting an idempotent into a projector and inclusion map is unique up to some reversible transformation relating two different bases for $S_{\Omega^\mathcal{A}}$.

The case of effects is handled in the same way, up to the caveat that we are thinking of all of the linear maps as acting contravariantly—that is, on the dual spaces, where the effects naturally live. That is, to split $\Pi_{\mathcal{E}^\mathcal{F}}$, one finds a vector space $S_{\mathcal{E}^\mathcal{A}}$ and a pair of linear maps

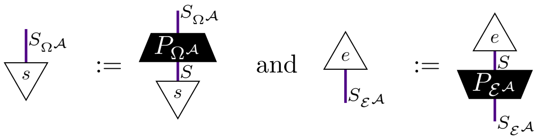

(15)

such that


(16)

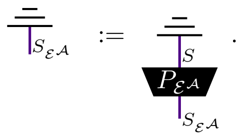
Here we find that $S^*_{\mathcal{E}^\mathcal{A}} \cong \mathsf{Span}[\mathcal{E}^\mathcal{F}]$ and so we think of this as the dual vector space of effects for the accessible GPT fragment. The map $P_{\mathcal{E}^\mathcal{A}}$ can then be though of as a projector mapping effects, viewed as covectors in the GPT, to effects, viewed as covectors in the accessible GPT fragment, and $I_{\mathcal{E}^\mathcal{A}}$ as an inclusion mapping effects, viewed as covectors in the accessible GPT fragment, to effects, viewed as covectors in the GPT. The choice of which idempotents to split, and which projector and inclusion map to split them into, then amounts to nothing more than picking a particular basis with which to represent the states and effects (one basis for the underlying GPT, and another for the accessible GPT fragment). That is, all of these choices simply result in equivalent accessible GPT fragments, as discussed in Ref. [14].

We can then define all of the components of the accessible GPT fragment in terms of the GPT fragment together with these projector and inclusion maps. In particular, we define the states and effects in the accessible GPT fragment as



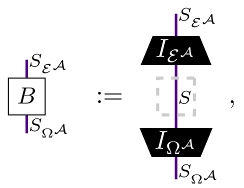
(17)

and so, in particular, the unit effect for the accessible GPT fragment is given by



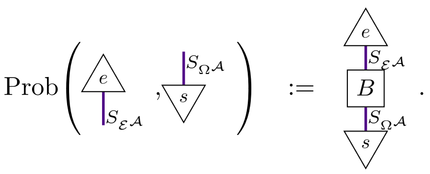
(18)

Finally, we define the probability rule for the accessible GPT fragment by including the states and effects in the accessible GPT fragment into the underlying GPT and computing the probability within the underlying GPT via Eq. (2). That is, we define a linear probability rule



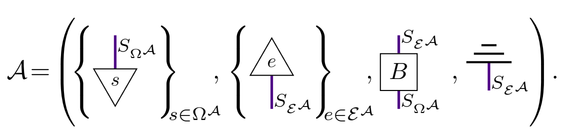
(19)

and use this to compute probabilities via



(20)

Succinctly, then, an accessible GPT fragment is specified by a quadruple



(21)

In this paper, we will assume that the number of extreme points in $\Omega^\mathcal{A}$ and $\mathcal{E}^\mathcal{A}$ are finite, as will be the case in any real experiment.

In short, from the GPT fragment characterising a given experiment, one can construct the associated accessible GPT fragment as follows. First, one closes $\Omega^\mathcal{F}$ and $\mathcal{E}^\mathcal{F}$ under classical processings. Then, one reconceptualizes the states and effects as living in their native subspaces, namely in $S_{\Omega^\mathcal{A}} \cong \mathsf{Span}[\Omega^\mathcal{F}]$ and $S_{\mathcal{E}^\mathcal{A}} \cong \mathsf{Span}[\mathcal{E}^\mathcal{F}]$ rather than in the underlying GPT's vector space S.

The notion of classical explainability for accessible GPT fragments is very closely related to that introduced above for GPTs and for GPT fragments. Since it is the main notion we use in this work, we discuss it in detail in the next section. It is straightforward to show (directly from the definitions and Eq. (12)) that classical explainability of a GPT fragment is equivalent to classical explainability of the associated accessible GPT fragment. This implies that one can work equally well with either the practical or the theoretical perspectives introduced above. Moreover, it implies that the particular choices that were made for the projection maps are irrelevant to the assessment of classicality (as one would expect).

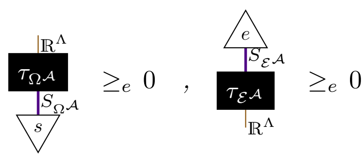
IV. Classical explainability of accessible GPT fragments

One can now ask whether a given accessible GPT fragment is classically-explainable. The appropriate notion of classical-explainability for accessible GPT fragments, first introduced in Ref. [14], is again a natural extension of the notion for standard GPTs:

Definition 4 (Simplicial-cone embeddings and simplex embeddings of an accessible GPT fragment)**.** *A* simplicial-cone embedding*, $\tau_\mathcal{A}$, of an accessible GPT fragment, $\mathcal{A}$, is defined by a set of ontic states Λ and a pair of linear maps*

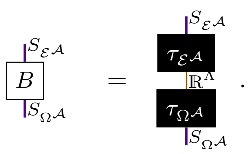
(22)

such that for all $s\in\Omega^\mathcal{A}$ and for all $e\in\mathcal{E}^\mathcal{A}$ we have



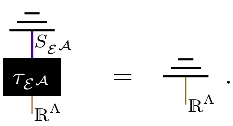
(23)

and such that



(24)

A simplicial-cone embedding is said to be a simplex embedding *if it moreover satisfies*

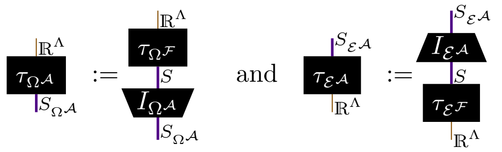


(25)

With these definitions in place it is straightforward to show an equivalence between classical-explainability of a GPT fragment and classical-explainability of the associated accessible GPT fragment.

Proposition 1. *Any simplex embedding for a GPT fragment $\mathcal{F}$ (Def. 3) can be converted into a simplex embedding for the associated accessible GPT fragment $\mathcal{A}$ (Def. 4), and vice versa.*

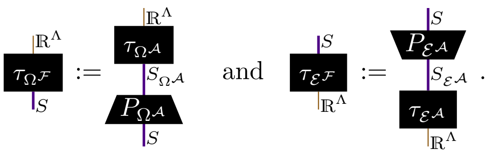
Proof. Given a simplex embedding $(\tau_{\Omega^{\mathcal{F}}}, \tau_{\mathcal{E}^{\mathcal{F}}})$ for the GPT fragment, one can construct a simplex embedding for the associated accessible GPT fragment by taking



(26)

The fact that this is a valid simplex embedding can be verified immediately from the definitions.

Similarly, given a simplex embedding $(\tau_{\Omega^{\mathcal{A}}}, \tau_{\mathcal{E}^{\mathcal{A}}})$ for the accessible GPT fragment, one can construct a simplex embedding for the GPT fragment by



(27)

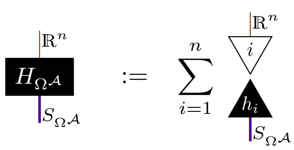
Again, the fact that this is a valid simplex embedding can be verified immediately from the definitions. □

Next, we show that the existence of a simplicial-cone embedding (or a simplex embedding) can be checked via a linear program; see Section C.

C. DERIVATION OF THE LINEAR PROGRAM

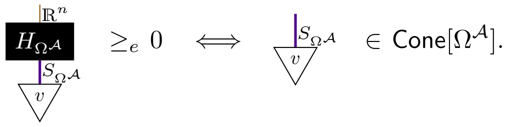
In this section we will show that, for any accessible GPT fragment, simplicial cone embeddability can be tested with a linear program. Note that the linear program discussed in the main text is simply the special case where the accessible GPT fragment lives inside quantum theory, that is, when $S_{\Omega^{\mathcal{A}}} \subseteq \mathsf{Herm}[\mathcal{H}]$ and $S_{\mathcal{E}^{\mathcal{A}}} \subseteq \mathsf{Herm}[\mathcal{H}]$ for some $\mathcal{H}$.

Let h_i be representative covectors for the extreme rays of the logical effect cone $\mathsf{Cone}[\Omega^{\mathcal{A}}]^*$, and let n be the number of these extreme rays. By definition, each of these h_i constitutes an inequality which defines a facet of the state cone $\mathsf{Cone}[\Omega^{\mathcal{A}}]$; conversely, every such facet is represented by some h_i. Then, one can define a map



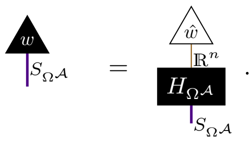
(28)

which takes any given state to the vector of n values which it obtains on these n facet inequalities. It follows that every vector v is in the state cone if and only if it is mapped by $H_{\Omega^{\mathcal{A}}}$ to a vector of positive values, i.e.,



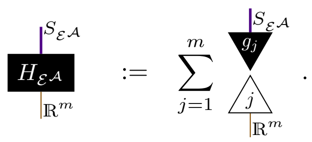
(29)

Furthermore, any valid inequality satisfied by all vectors in the logical effect cone can be written as a positive linear sum of facet inequalities. Equivalently, one has that for any $w \in \mathsf{Cone}[\Omega]^*$, there exists $\hat{w} \geq_e 0$ such that



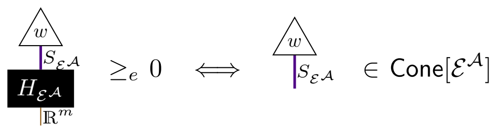
(30)

Similarly, let g_j be representative vectors for the extreme rays of $\mathsf{Cone}[\mathcal{E}^{\mathcal{A}}]^*$. Let the number of such extreme rays be m, and define



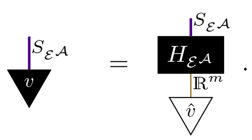
(31)

Then one has that



(32)

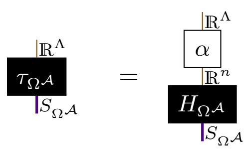
and that for any $v \in \mathsf{Cone}[\mathcal{E}^{\mathcal{A}}]^*$, there exists $\hat{v} \geq_e 0$ such that



(33)

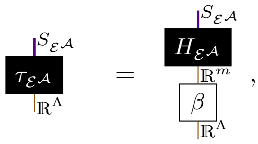
Recall that a simplicial-cone embedding is defined in terms of linear maps $\tau_{\Omega^{\mathcal{A}}}$ and $\tau_{\mathcal{E}^{\mathcal{A}}}$. We now prove a useful lemma relating $H_{\Omega^{\mathcal{A}}}$ with $\tau_{\Omega^{\mathcal{A}}}$ and $H_{\mathcal{E}^{\mathcal{A}}}$ with $\tau_{\mathcal{E}^{\mathcal{A}}}$.

Lemma 5. *In any simplicial-cone embedding defined by linear maps $\tau_{\Omega^{\mathcal{A}}}$ and $\tau_{\mathcal{E}^{\mathcal{A}}}$, the map $\tau_{\Omega^{\mathcal{A}}}$ factors through $H_{\Omega^{\mathcal{A}}}$ as*



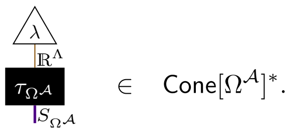
(34)

and $\tau_{\mathcal{E}^{\mathcal{A}}}$ factors through $H_{\mathcal{E}^{\mathcal{A}}}$ as


(35)

where $\alpha:\mathbb{R}^n \to \mathbb{R}^\Lambda$ and $\beta:\mathbb{R}^\Lambda\to\mathbb{R}^m$ are matrices with nonnegative entries.

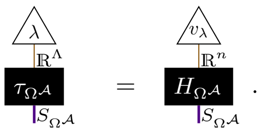
Proof. Since $\tau_{\Omega^{\mathcal{A}}}$ maps vectors in the state cone to points in the simplicial cone, it follows that for all $\lambda\in\Lambda$, one has



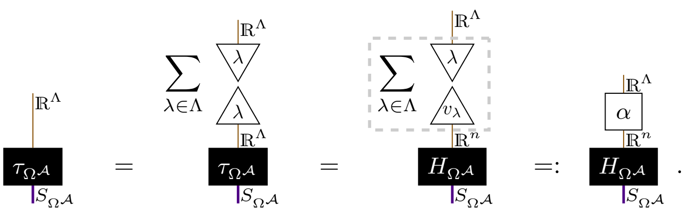
(36)

To see that this is the indeed the case, note that Eq. (36) asserts that the process on its LHS is in the dual cone—i.e., it evaluates to a non-negative number on arbitrary vectors in the state cone. This is indeed the case, because if one composes an arbitrary vector in the state cone with $\tau_{\Omega^{\mathcal{A}}}$, one gets a vector in the simplicial cone by assumption; then, the effect λ simply picks out the (necessarily non-negative) relevant coefficient corresponding to the λ basis element.

Hence, Eq. (30) implies that there exists a non-negative covector v_λ such that


(37)

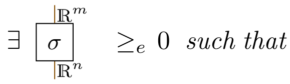
Now, one simply inserts a resolution of the identity on the system coming out of $\tau_{\Omega^{\mathcal{A}}}$ and uses the above result to obtain the desired factorisation:


(38)

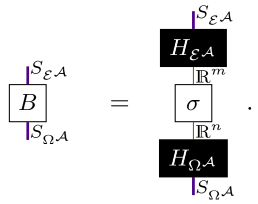
The proof for the factorisation of $\tau_{\mathcal{E}^{\mathcal{A}}}$ is almost identical, except that one inserts a resolution of the identity for the ingoing system rather than the outgoing system. □

With this in place, our main theorem is simple to prove. Linear Program 1 in the main text is a special case of this where the underlying GPT is taken to be quantum.

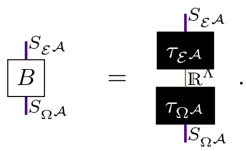
Theorem 6. *Consider any accessible GPT fragment $\mathcal{A}:=\{\Omega^{\mathcal{A}},\mathcal{E}^{\mathcal{A}},B,\mathbf{u}\}$ with state cone characterized by a matrix $H_{\Omega^{\mathcal{A}}}$ (whose codomain is dimension n) and effect cone characterized by matrix $H_{\mathcal{E}^{\mathcal{A}}}$ (whose domain is dimension m). Then the accessible GPT fragment $\mathcal{A}$ is classically explainable if and only if*



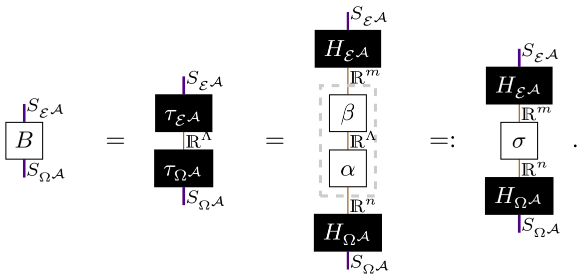
(39)


(40)

Proof. A simplicial-cone embedding is given by a $\tau_{\mathcal{E}^{\mathcal{A}}}$ and $\tau_{\Omega^{\mathcal{A}}}$ satisfying

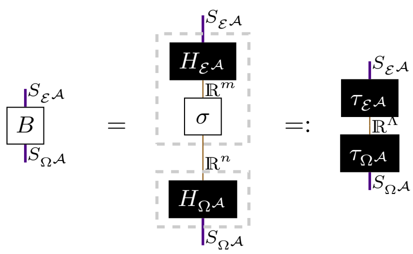

(41)

If these exist, one can apply Lemma 5 to write


(42)

Hence, we arrive at a decomposition of the form of Eq. (40), where furthermore $\sigma\geq_e 0$, since α and β are both entry-wise positive.

Conversely, if there is a decomposition of the form given by Eq. (40), then we can define $\tau_{\mathcal{E}^{\mathcal{A}}}$ and $\tau_{\Omega^{\mathcal{A}}}$ as

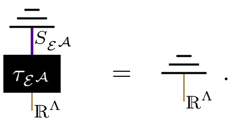

(43)

to yield a valid simplicial-cone embedding. In particular, (i) $\tau_{\mathcal{E}^{\mathcal{A}}}$ and $\tau_{\Omega^{\mathcal{A}}}$ are clearly linear; (ii) $\tau_{\Omega^{\mathcal{A}}}$ maps states into the simplicial cone, since $H_{\Omega^{\mathcal{A}}}$ satisfies Eq. (29); and (iii) $\tau_{\mathcal{E}^{\mathcal{A}}}$ maps effects into the dual of the simplicial cone, since $H_{\mathcal{E}^{\mathcal{A}}}$ satisfies Eq. (32) and $\sigma\geq_e 0$. □

The condition expressed in the statement of Theorem 6 provides us with our linear program for testing for classical-explainability of an accessible GPT fragment, or equivalently, classical-explainability of a GPT fragment from which it came. The core of the linear program is finding a suitable matrix $\sigma \geq_e 0$. From a solution σ, one can construct a simplicial-cone embedding via Eq. (43). From this, one can construct a simplex embedding, which we do in the next section. In the section after that, we explicitly construct the ontological model for the GPT which is equivalent to the simplex-embedding.

I. From simplicial-cone embeddings to simplex embeddings

In Ref. [14], we showed that if a simplicial-cone embedding exists then so too does a simplex embedding. Recall that the latter is just a simplicial-cone embedding satisfying an additional constraint on $\tau_{\mathcal{EA}}$, namely, that:

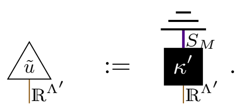


(44)

We will now show how, given any simplicial-cone embedding given by $\tau'_{\Omega\mathcal{A}}$ and $\tau'_{\mathcal{EA}}$ and ontic state space Λ', we can construct a simplex embedding. This construction is useful because, by the results of Ref. [14], it is equivalent to constructing an ontological model for the GPT fragment.

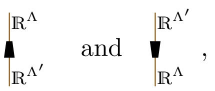
The construction essentially removes superfluous ontic states and then rescales the maps $\tau'_{\Omega\mathcal{A}}$ and $\tau'_{\mathcal{EA}}$ (in a manner that can depend on the ontic state) to ensure that the representation of the ignoring operation is given by the all-ones vector, as Eq. (44) states. In what follows, we explain how this works.

First, let us define $\Lambda := \mathsf{Supp}[\tilde{u}] \subseteq \Lambda'$—this will be the ontic state space for the simplex embedding—where $\tilde{u}$ is defined as



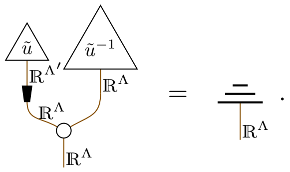
(45)

We can define a projection and an inclusion map for this subspace, which we denote as:



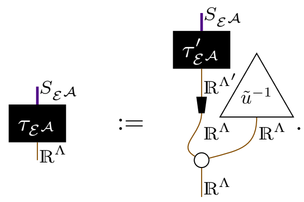
(46)

respectively. Within this subspace, we can define an inverse of $\tilde{u}$, which we denote by $\tilde{u}^{-1}$, as the covector that satisfies:



(47)

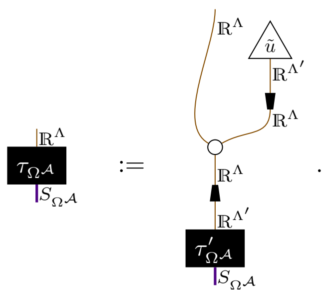
With these, we define the map $\tau_{\mathcal{EA}}$ (which describes the embedding of the effects) as



(48)

This is just the required removal of ontic states (as dictated by the inclusion map) rescaling of $\tau'_{\mathcal{EA}}$ by the appropriate real values (as dictated by $\tilde{u}^{-1}$). These values are chosen to ensure that Eq. (44) is satisfied (as one can easily verify, as a direct consequence of Eq. (47)) and to ensure that it maps effects to entrywise-positive covectors. This is an explicit description of the rescaling matrix R discussed in the main text.

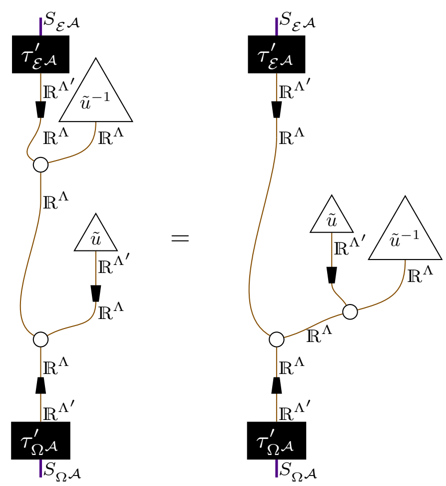
Then, we can define the map $\tau_{\Omega\mathcal{A}}$ (that describes the simplex embedding of the states) as:



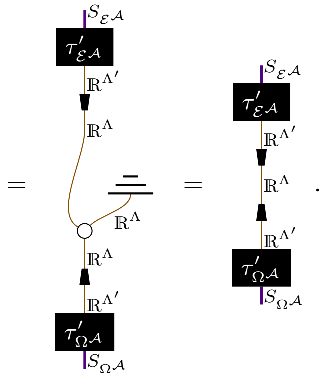
(49)

It is simple to verify that $\tau_{\Omega\mathcal{A}}$ maps states to entrywise-positive vectors. All that then remains to be shown is that when $\tau_{\Omega\mathcal{A}}$ and $\tau_{\mathcal{EA}}$ are composed that we reproduce the probability rule B.

To see this, first note that



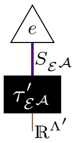
(50)



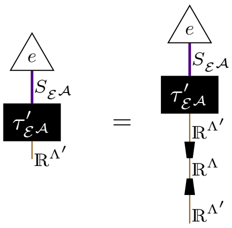
(51)

Notice that the right-hand-side of Eq. (51) would be precisely equal to B if it were not for the projection and inclusion maps in between $\tau_{\Omega^{\mathcal{A}}}$ and $\tau_{\mathcal{E}^{\mathcal{A}}}$. Hence, as a final step we need to show that these maps are redundant in this expression.

To see this, first recall that for all $e\in\mathcal{E}^{\mathcal{A}}$, there exists $e^{\perp}\in\mathcal{E}^{\mathcal{A}}$ such that $e+e^{\perp}=\bar{\top}_{\mathcal{E}^{\mathcal{A}}}$. Since e and $e^{\perp}$ are both mapped to entrywise nonnegative covectors by $\tau'_{\mathcal{E}^{\mathcal{A}}}$, and because the sum of these covectors must be the vector $\tilde{u}$, it follows that

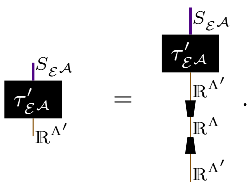
 $\in \mathsf{Supp}[\tilde{u}]=\Lambda \qquad \forall e\in\mathcal{E}^{\mathcal{A}}.$ (52)

Hence,



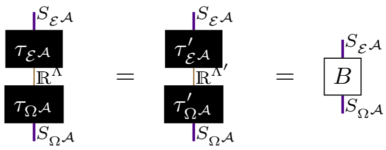
(53)

for all $e\in\mathcal{E}^{\mathcal{A}}$. As $\mathcal{E}^{\mathcal{A}}$ spans $S_{\mathcal{E}^{\mathcal{A}}}$ we therefore have that:



(54)

Putting this all together, we therefore have that

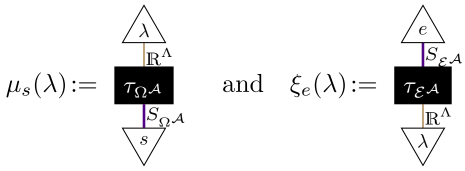


(55)

which completes the result.

II. From simplex embeddings to ontological models

We have therefore seen how to transform a solution (σ) to the linear program into a simplicial-cone embedding, and from that to a simplex embedding. An explicit ontological model can be stated directly in terms of this embedding. Specifically, one defines the epistemic states and response functions in the ontological model as



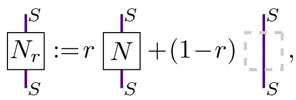
(56)

for all $\lambda\in\Lambda$, $s\in\Omega^{\mathcal{A}}$, $e\in\mathcal{E}^{\mathcal{A}}$. That this is a valid ontological model follows from the results of Refs. [5, 14].

D. AN OPERATIONAL MEASURE OF NONCLASSICALITY

Thus far, we have only discussed the qualitative question of *whether or not* a classical explanation exists for a given scenario. A natural next question is to introduce quantitative measures of the *degree* of nonclassicality in one's scenario. A particularly useful approach to doing this would be to introduce a resource theory [24] of generalized noncontextuality and finding monotones [25] therein. However, such an approach has not yet been developed. Therefore, here we take an approach motivated by the fact that *every* experiment admits of a classical explanation when subject to sufficient depolarizing noise [26, 27]. Hence, our approach is to quantify the robustness of one's nonclassicality—that is, the amount of noise which must be applied to one's data until it admits of a classical explanation. This is by no means a uniquely privileged measure, but it is operationally well-motivated.

There are many reasonable noise models, and which of these is most suitable depends on one's physical scenario. In this section, we show how one can adapt our linear program to quantify robustness of nonclassicality with respect to any noise model which treats noise as the probabilistic application of a channel to all states in the experiment. That is, we consider arbitrary noise models of the form



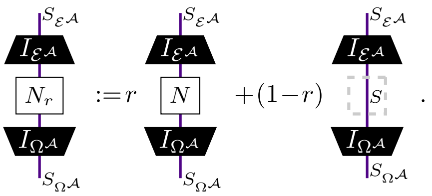
(57)

where N is an arbitrary channel (representing the noise). Note that Eq. (57) describes the noise as a channel in the full underlying GPT space S. One could alternatively describe noise as a channel acting on the spaces in which accessible GPT fragment lives (namely, a channel from $S_{\Omega^{\mathcal{A}}}$ to $S_{\mathcal{E}^{\mathcal{A}}}$); this is simply a further freedom in one's choice of noise model, and the techniques of this section apply to either approach.

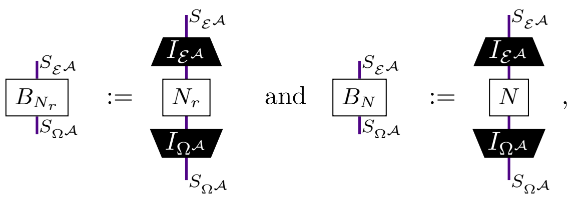
Each different noise model leads to a distinct measure of robustness of nonclassicality. Perhaps the most common quantum noise models are those of this form and where N is chosen to be either the completely depolarizing channel or the completely dephasing channel in a particular basis. For concreteness, in the quantum case, our open-source code is implemented assuming depolarizing noise, as we detail below. For arbitrary GPTs, however, there is not necessarily a unique, well-defined maximally mixed state, and so the completely depolarising channel is not necessarily well-defined. In the GPT case, our open-source code therefore asks the user to specify a state to act as the maximally mixed state, which is then used to construct a completely depolarising channel, and the robustness to this noise channel is then computed and output by the program. We explore the case of dephasing noise in Ref. [28].

Suppose the linear program discussed in Theorem 6 determines that a particular accessible GPT fragment does not admit of a simplex embedding. The natural next question tackled in this section is then rephrased as how much noise must one's experiment be subject to until it becomes simplex-embeddable. That is, what is the minimum value of r for which one's experiment becomes classically-explainable?

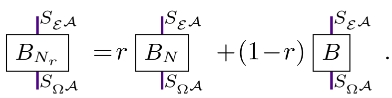
To address this question, we first translate Eq. (57) into its description at the level of the accessible GPT fragment, simply by applying the appropriate inclusion maps (and applying linearity of the inclusion map on the RHS):


(58)

Then, we define


(59)

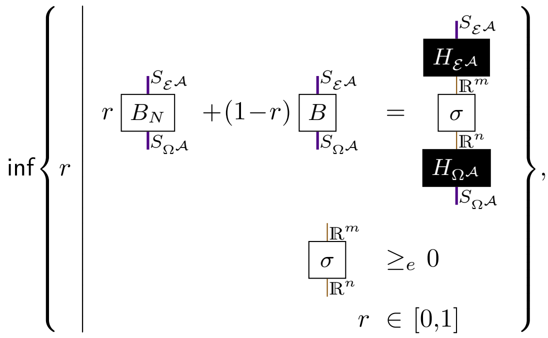
to write


(60)

Hence, we see that the effect of the noise is simply to modify the linear map which captures the probability rule.

Clearly then, for a particular value of r, we can ask whether the new accessible GPT fragment, defined by replacing the old probability rule B with this new one B_{N_r}, is simplex embeddable, simply by running the linear program. However, what is more interesting is to allow for r to be an additional variable, and to ask: what is the minimal value of r such that the associated accessible GPT fragment is simplex embeddable?

This is formulated as the following optimization problem:

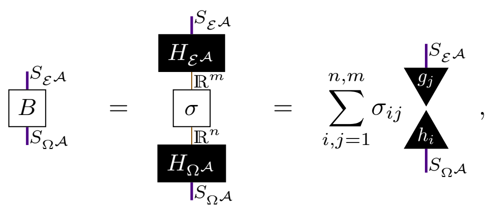

(61)

which is also a linear program, since the only unknown quantities are the elements of σ and r.

This linear program tells us the minimal amount of the noise channel N which needs to be added to the experiment in order that it admit of a classical explanation. Similarly to before, the particular σ that is found for the minimal value of r, can then be used to construct an explicit $\tau_{\Omega\mathcal{A}}$ and $\tau_{\mathcal{E}\mathcal{A}}$ which define the simplex embedding of the accessible GPT fragment that results after this amount of noise is applied.

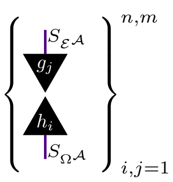
E. BOUNDING THE NUMBER OF ONTIC STATES

Let us assume that an accessible GPT fragment $\mathcal{G}$ satisfies Theorem 6, and hence we have a decomposition of the linear map B as

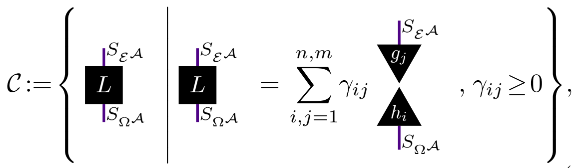

(62)

where $\sigma_{ij} \geq 0$.

It follows that B belongs to a particular convex cone $\mathcal{C} \subset \mathcal{L}(S_{\Omega\mathcal{A}}, S_{\mathcal{E}\mathcal{A}})$ living inside the real vector space of linear maps from $S_{\Omega\mathcal{A}}$ to $S_{\mathcal{E}\mathcal{A}}$; namely, the convex cone $\mathcal{C}$ given by the conic closure of a particular set of linear maps:


(63)

In other words, B belongs to the cone

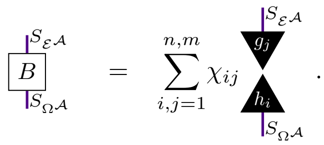


(64)

which is clear from Eq. (62).

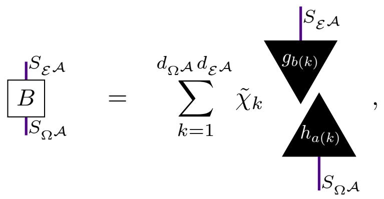
We can therefore apply Carathéodory's theorem which, in this context, states that any linear map living inside the cone $\mathcal{C}$ can be decomposed as a conic combination of at most $\mathsf{dim}[\mathcal{L}(S_{\Omega\mathcal{A}},S_{\mathcal{E}\mathcal{A}})]=\mathsf{dim}[S_{\Omega\mathcal{A}}]\mathsf{dim}[S_{\mathcal{E}\mathcal{A}}]=:d_{\Omega\mathcal{A}}d_{\mathcal{E}\mathcal{A}}$ vertices of $\mathcal{C}$.

Now, as we know that $B\in\mathcal{C}$, this means that there should exist coefficients $\chi_{ij}\geq 0$ such that the number of $\chi_{ij}\neq 0$ is at most $d_{\Omega\mathcal{A}}d_{\mathcal{E}\mathcal{A}}$. We write this new decomposition as



(65)

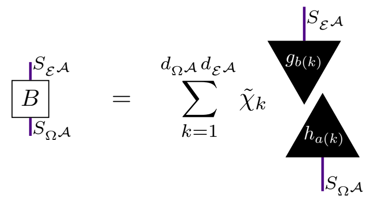
To make explicit that this decomposition only uses at most $d_{\Omega\mathcal{A}}d_{\mathcal{E}\mathcal{A}}$ non-zero elements, we switch to a single index $k\in\{1,...,d_{\Omega\mathcal{A}}d_{\mathcal{E}\mathcal{A}}\}=:K$ and write it as



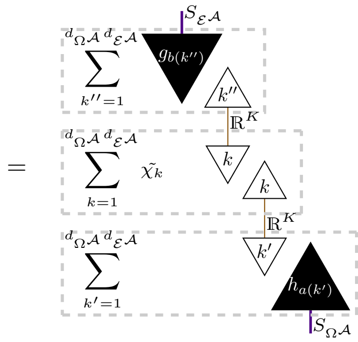
(66)

where a and b are functions such that $\tilde{\chi}_k=\chi_{a(k)b(k)}$.

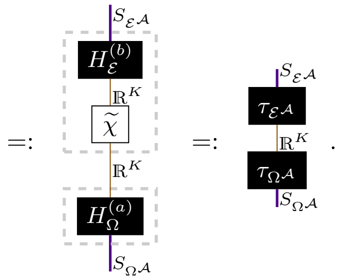
Now it is simple to rewrite this into the form of a simplex embedding with ontic states indexed by $k\in K$:



(67)



(68)



(69)

Thus, we have found a simplex embedding for the accessible GPT fragment into a simplex with vertices labeled by K. Consequently, the maximum number of ontic states that need be considered is $\mathsf{dim}[\mathcal{L}(S_{\Omega\mathcal{A}},S_{\mathcal{E}\mathcal{A}})]=\mathsf{dim}[S_{\Omega\mathcal{A}}]\mathsf{dim}[S_{\mathcal{E}\mathcal{A}}]=d_{\Omega\mathcal{A}}d_{\mathcal{E}\mathcal{A}}$. In the case of standard GPTs, where $S_{\Omega\mathcal{A}}=S_{\mathcal{E}\mathcal{A}}$ and $d_{\Omega\mathcal{A}}=d_{\mathcal{E}\mathcal{A}}=d$, the GPT dimension, we find that the maximal number of ontic states that need be considered is given by d^2. This is the same bound that was discovered in Ref. [2].

It is unclear whether a tighter bound can be found. Certainly, in some specific examples, less than d^2 ontic states are required. Given a solution to the linear program, as a matrix σ, one can find a (potentially tighter) upper bound on the number of ontic states by finding the nonnegative rank of σ. The nonnegative rank is the minimal dimension of the vector space through which one can factor σ in such a way that the two factors are nonnegative. It is easy to see that this defines an ontological model with a number of ontic states equal to the nonnegative rank, as the two factors of σ can be used to define the $\tau_{\Omega\mathcal{A}}$ and $\tau_{\mathcal{E}\mathcal{A}}$ for a simplex embedding. There may, however, be many different σ which are valid solutions to the linear program, and moreover, we do not have any proof that these necessarily have the same nonnegative rank. In order to find the minimal number of ontic states, one may therefore have to minimise the nonnegative rank over all possible solutions to the linear program. It is not clear whether there is an efficient method to solve this optimisation problem.

F. WORKED EXAMPLES

In this section, we analyze the three illustrative examples we introduced in the main text as well as one further example from the GPT known as Boxworld [21]. We start with finite sets of states Ω and effects $\mathcal{E}$, and we show in detail how our linear program techniques can be used to assess their classicality. In Section A of this Supplemental Material, we use these examples to comment on the related works in Refs. [2, 4].

I. Example 1

Consider the set of four quantum states

$$\Omega=\{|0\rangle\langle 0|,|1\rangle\langle 1|,|+\rangle\langle +|,|-\rangle\langle -|\},\qquad(70)$$

on a qubit. Consider moreover the finite set of effects

$$\mathcal{E} = \{|0\rangle\langle 0|, |1\rangle\langle 1|, |+\rangle\langle +|, |-\rangle\langle -|, \mathbb{1}, 0\}. \tag{71}$$

Now the question is: *are the statistics obtained by composing any state-effect pair classically-explainable?* To answer this question in the affirmative, we now show that these sets of states and effects satisfy with the requirements of Linear Program 1.

First, take the set of Hermitian operators $O := \left\{\frac{1}{\sqrt{2}}\mathbb{1}, \frac{1}{\sqrt{2}}\sigma_x, \frac{1}{\sqrt{2}}\sigma_z\right\}$, where σ_x and σ_z are the Pauli-X and Pauli-Z operators, respectively. Notice that O is an orthonormal basis for the subspace S_Ω, and that in this basis the states in Ω are expressed as:

$$\begin{aligned} |0\rangle\langle 0| &\leftrightarrow \tfrac{1}{\sqrt{2}}[1,0,1]^T, \\ |1\rangle\langle 1| &\leftrightarrow \tfrac{1}{\sqrt{2}}[1,0,-1]^T, \\ |+\rangle\langle +| &\leftrightarrow \tfrac{1}{\sqrt{2}}[1,1,0]^T, \\ |-\rangle\langle -| &\leftrightarrow \tfrac{1}{\sqrt{2}}[1,-1,0]^T. \end{aligned}$$

In this coordinate system, the Hermitian operators corresponding to the facet inequalities for $\mathsf{Cone}[\Omega]$ (which has 4 facets) expressed in the basis O read:

$$\begin{aligned} h_1^\Omega &= [1,1,1]^T, \\ h_2^\Omega &= [1,1,-1]^T, \\ h_3^\Omega &= [1,-1,1]^T, \\ h_4^\Omega &= [1,-1,-1]^T. \end{aligned}$$

Using these, we build the linear map H_Ω that maps elements of S_Ω to vectors in $\mathbb{R}^4$:

$$H_\Omega = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}. \tag{72}$$

For instance,

$$H_\Omega[|0\rangle\langle 0|] = [\sqrt{2}, 0, \sqrt{2}, 0]^T. \tag{73}$$

Next, let us discuss the inclusion map for the case of states, I_Ω. In this example we are working with a qubit system (that is, a quantum system in $\mathcal{H}_2$), hence I_Ω will map S_Ω into $\mathsf{Herm}[\mathcal{H}_2]$. Now, notice that our chosen basis for S_Ω is $O = \{\frac{1}{\sqrt{2}}\mathbb{1}, \frac{1}{\sqrt{2}}\sigma_x, \frac{1}{\sqrt{2}}\sigma_z\}$, while a basis for $\mathsf{Herm}[\mathcal{H}_2]$ is $\{\frac{1}{\sqrt{2}}\mathbb{1}, \frac{1}{\sqrt{2}}\sigma_x, \frac{1}{\sqrt{2}}\sigma_y, \frac{1}{\sqrt{2}}\sigma_z\}$, where σ_y is the Pauli-Y operator. Hence, the matrix representation of I_Ω in these bases is

$$I_\Omega = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{74}$$

Similarly, one can define the corresponding bases, facet inequality operators, $H_\mathcal{E}$, and inclusion map for the set of effects $\mathcal{E}$. In a nutshell, notice that O is also a basis for the subspace $S_\mathcal{E}$, hence:

$$\begin{aligned} 0 &\leftrightarrow [0,0,0]^T, \\ |0\rangle\langle 0| &\leftrightarrow \tfrac{1}{\sqrt{2}}[1,0,1]^T, \\ |1\rangle\langle 1| &\leftrightarrow \tfrac{1}{\sqrt{2}}[1,0,-1]^T, \\ |+\rangle\langle +| &\leftrightarrow \tfrac{1}{\sqrt{2}}[1,1,0]^T, \\ |-\rangle\langle -| &\leftrightarrow \tfrac{1}{\sqrt{2}}[1,-1,0]^T, \\ \mathbb{1} &\leftrightarrow [\sqrt{2},0,0]^T. \end{aligned}$$

The facet-defining inequalities for $\mathsf{Cone}[\mathcal{E}]$ are hence the same as for $\mathsf{Cone}[\Omega]$. Using $\{h_1^\Omega, h_2^\Omega, h_3^\Omega, h_4^\Omega\}$ to define the linear map $H_\mathcal{E}$ that maps elements of $S_\mathcal{E}$ to vectors in $\mathbb{R}^4$, one obtains

$$H_\mathcal{E} = H_\Omega. \tag{75}$$

Finally, the inclusion map $I_\mathcal{E}$ embeds $S_\mathcal{E}$ into $\mathsf{Herm}[\mathcal{H}_2]$. Following the same argument as for the case of states, one hence obtains

$$I_\mathcal{E} = I_\Omega. \tag{76}$$

Now that we have explicit forms for H_Ω, I_Ω, $H_\mathcal{E}$, and $I_\mathcal{E}$, we can use Linear Program 1 to assess the classical-explainability of the data generated by $(\Omega, \mathcal{E})$. More precisely, the statistics obtained by composing any state-effect pair is classically-explainable if there exists a matrix σ with non-negative entries such that

$$I_\mathcal{E}^T \cdot I_\Omega = H_\mathcal{E}^T \cdot \sigma \cdot H_\Omega. \tag{77}$$

One can then check that the following matrix does the job:

$$\sigma = \frac{1}{4}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \frac{1}{4}\mathbb{1}_4. \tag{78}$$

The simplex embedding is found from σ and the H matrices (following Eq. (43) and Section C I) to be given by

$$\tau_\Omega = \tfrac{1}{2}\tau_\mathcal{E} \quad \text{and} \quad \tau_\mathcal{E} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}. \tag{79}$$

From this, one can find an ontological model for the scenario (following Section C II). The model has four ontic states, over which the epistemic states for the four quantum states in Eq. (70) are

$$[\tfrac{1}{2},0,\tfrac{1}{2},0]^T, \quad [0,\tfrac{1}{2},0,\tfrac{1}{2}]^T, \quad [\tfrac{1}{2},\tfrac{1}{2},0,0]^T, \quad [0,0,\tfrac{1}{2},\tfrac{1}{2}]^T, \tag{80}$$

and the response functions for the six effects in Eq. (71) are

$$[1,0,1,0]^T, \quad [0,1,0,1]^T, \quad [1,1,0,0]^T, \quad [0,0,1,1]^T, \quad [1,1,1,1]^T, \quad [0,0,0,0]^T, \tag{81}$$

respectively. One can directly check that this model reproduces the quantum statistics, and that it corresponds to the model in Fig. 1 of the main text.

II. Example 2

Here we consider the example of a quantum system of dimension four. The sets of states and effects that we consider are, respectively,

$$\Omega = \{|0\rangle\langle 0|, |1\rangle\langle 1|, |2\rangle\langle 2|, |3\rangle\langle 3|\}, \tag{82}$$

and

$$\mathcal{E} = \{|0\rangle\langle 0| + |1\rangle\langle 1|, |1\rangle\langle 1| + |2\rangle\langle 2|, |2\rangle\langle 2| + |3\rangle\langle 3|, |3\rangle\langle 3| + |0\rangle\langle 0|, \mathbb{1}_4, 0\}. \tag{83}$$

Notice that although $\mathsf{Herm}[\mathcal{H}_4]$ is a 16-dimensional space, S_Ω is 4-dimensional whilst $S_\mathcal{E}$ is only 3-dimensional.

Now, choose

$$\{\tfrac{1}{2}\mathbb{1}_4, \tfrac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1| - |2\rangle\langle 2| - |3\rangle\langle 3|),$$
$$\tfrac{1}{2}(-|0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2| - |3\rangle\langle 3|),$$
$$\tfrac{1}{2}(|0\rangle\langle 0| - |1\rangle\langle 1| + |2\rangle\langle 2| - |3\rangle\langle 3|)\},$$

as an othonormal basis of Hermitian operators for S_Ω. Notice that in this representation, the four states that define Ω are associated with 4 vertices of a cube, and define a tetrahedron. Since the facets of $\mathsf{Cone}[\Omega]$ correspond to the facets of such tetrahedron, the number of facet-defining inequalities (i.e., rows of H_Ω) is 4.

In addition, consider the following subbasis:

$$\{\tfrac{1}{2}\mathbb{1}_4, \tfrac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1| - |2\rangle\langle 2| - |3\rangle\langle 3|),$$
$$\tfrac{1}{2}(-|0\rangle\langle 0| + |1\rangle\langle 1| + |2\rangle\langle 2| - |3\rangle\langle 3|)\}.$$

Linear combinations of these three Hermitian operators can actually yield the vertices of $\mathcal{E}$, and hence we can take them as a basis of Hermitian operators for $S_\mathcal{E}$ which is hence of dimension 3. Notice that in this representation, the four effects that define $\mathcal{E}$ correspond to the four vertices of a square, and hence the number of facets of $\mathsf{Cone}[\mathcal{E}]$ is 4.

Similarly to the previous example, one can then compute the linear maps H_Ω and $H_\mathcal{E}$, and here they are then found to be

$$H_\Omega = \begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}, \quad H_\mathcal{E} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}. \tag{84}$$

On the other hand, notice that $\mathsf{Herm}[\mathcal{H}_4]$ is spanned by a Hermitian operator basis of 16 elements. The inclusion maps I_Ω and $I_\mathcal{E}$ then read

$$I_\Omega = \begin{bmatrix} \mathbb{1}_4 \\ \mathbf{0}_{12\times 4} \end{bmatrix}, \quad I_\mathcal{E} = \begin{bmatrix} \mathbb{1}_3 \\ \mathbf{0}_{13\times 3} \end{bmatrix}, \tag{85}$$

where the matrix $\mathbf{0}_{a\times b}$ has dimension $a\times b$ and all entries equal to 0.

Now the question is whether the statistics from every pair of state-effect drawn from $(\Omega,\mathcal{E})$ can be explained classically. By taking the matrix

$$\sigma = \tfrac{1}{4}\mathbb{1}_4, \tag{86}$$

one can straightforwardly check that $(\Omega,\mathcal{E})$ satisfy the condition of Linear Program 1, answering the question in the affirmative.

The simplex embedding is found from σ and the Hs (following Eq. (43) and Section C I) to be given by

$$\tau_\Omega = \tfrac{1}{2}\begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & -1 \end{bmatrix}, \quad \text{and} \quad \tau_\mathcal{E} = \tfrac{1}{2}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix}. \tag{87}$$

From this, one can find an ontological model for the scenario (following Section C II). The model has four ontic states, over which the epistemic states for the four quantum states in Eq. (82) are

$$[0,1,0,0]^T, \quad [1,0,0,0]^T, \quad [0,0,1,0]^T, \quad [0,0,0,1]^T, \tag{88}$$

and the response functions for the six effects in Eq. (83) are

$$[1,1,0,0]^T, \quad [1,0,1,0]^T, \quad [0,0,1,1]^T, \quad [0,1,0,1]^T, \quad [1,1,1,1]^T, \quad [0,0,0,0]^T, \tag{89}$$

respectively. One can directly check that this model reproduces the quantum statistics, and that it corresponds to the model in Fig. 2 of the main text.

III. Example 3

Our third example has the same states and effects as our first example, except that the effects have been rotated by about the σ_y axis by $\frac{\pi}{4}$. Explicitly, we consider the states

$$\Omega = \{|0\rangle\langle 0|, |1\rangle\langle 1|, |+\rangle\langle +|, |-\rangle\langle -|\}, \tag{90}$$

and effects

$$\mathcal{E} = \{\mathcal{R}(|0\rangle\langle 0|), \mathcal{R}(|1\rangle\langle 1|), \mathcal{R}(|+\rangle\langle +|), \mathcal{R}(|-\rangle\langle -|), \mathbb{1}, 0\} \tag{91}$$

where $\mathcal{R}(\_) := R_y(\frac{\pi}{4})(\_)R_y(-\frac{\pi}{4})$ and $R_y(\theta) := \begin{pmatrix} \cos(\frac{\theta}{2}) & \sin(\frac{\theta}{2}) \\ -\sin(\frac{\theta}{2}) & \cos(\frac{\theta}{2}) \end{pmatrix}$. This keeps us in the $\sigma_x - \sigma_z$ plane of the Bloch ball, and hence we can work with the same operator basis as in the first example. As before, then, we therefore obtain

$$H_\Omega = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \tag{92}$$

and

$$I_\Omega = I_\mathcal{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{93}$$

Unlike in the first example, however, the rotated effects lead to a different $H_{\mathcal{E}}$, namely

$$H_{\mathcal{E}}=\begin{bmatrix}13860 & 0 & 19601\\ 13860 & 0 & -19601\\ 13860 & 19601 & 0\\ 13860 & -19601 & 0\end{bmatrix}. \tag{94}$$

Note that the code doesn't work with irrational numbers directly, so the above $H_{\mathcal{E}}$ is a rational approximation which works well for our purposes.

Now that we have explicit forms for H_Ω, I_Ω, $H_{\mathcal{E}}$, and $I_{\mathcal{E}}$, we can use Linear Program 1 to assess the classical-explainability of the states and effects.

We find that this set of states and effects is *not* simplex embeddable, and hence is not classically explainable. Indeed, this remains the case until we depolarise by $r=\frac{8119}{27720}\sim 1-\frac{1}{\sqrt{2}}$. Once we have depolarised by this amount, then we find that Linear Program 1 is satisfiable for

$$\sigma=\frac{1}{110880}\begin{bmatrix}1&0&1&0\\0&1&0&1\\1&1&0&0\\0&0&1&1\end{bmatrix}. \tag{95}$$

The simplex embedding for the depolarised scenario can then be computed from σ and the H matrices (following Eq. (43) and Section C I) to be given by

$$\tau_{\Omega\mathcal{A}}=\frac{1}{2\sqrt{2}}\begin{bmatrix}1&1&1\\1&1&-1\\1&-1&1\\1&-1&-1\end{bmatrix} \quad\text{and}\quad \tau_{\mathcal{E}\mathcal{A}}=\frac{1}{\sqrt{2}}\begin{bmatrix}1&\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\1&\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\\1&-\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{2}}\\1&-\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{2}}\end{bmatrix}, \tag{96}$$

From these, one can then find an ontological model for the depolarised scenario (following Section C II).

Specifically, we find that the states in the depolarised scenario are represented as

$$[\tfrac{1}{2},0,\tfrac{1}{2},0]^T,\quad [0,\tfrac{1}{2},0,\tfrac{1}{2}]^T,\quad [\tfrac{1}{2},\tfrac{1}{2},0,0]^T,\quad [0,0,\tfrac{1}{2},\tfrac{1}{2}]^T, \tag{97}$$

and the response functions for the six effects are

$$[1,\tfrac{1}{2},\tfrac{1}{2},0]^T,\quad [0,\tfrac{1}{2},\tfrac{1}{2},1]^T,\quad [\tfrac{1}{2},0,1,\tfrac{1}{2}]^T,\quad [\tfrac{1}{2},1,0,\tfrac{1}{2}]^T,$$
$$[1,1,1,1]^T,\quad [0,0,0,0]^T. \tag{98}$$

One can then directly check that this model reproduces the quantum statistics in the scenario depolarised by $r=1-\frac{1}{\sqrt{2}}$, and that it corresponds to the model in Fig. 3 of the main text.

IV. Example 4

This final example is quite different to the ones we have so far presented in the sense that it does not pertain to quantum states and effects. Instead, here we will apply our technique to assess the classicality of experiments performed with states and effects for systems in the generalised probabilistic theory known as Boxworld [29].

Consider a single Boxworld system. Now, take the finite set of states

$$\Omega=\{[1,1,0]^T,[1,0,1]^T,[1,-1,0]^T,[1,0,-1]^T\}. \tag{99}$$

Given the states we started with, it follows that $\mathsf{ConvHull}[\Omega]$ corresponds to the full state space of the Boxworld system.

Similarly, consider the following finite set of effects

$$\mathcal{E}=\{\tfrac{1}{2}[1,-1,-1]^T,\tfrac{1}{2}[1,1,-1]^T,\tfrac{1}{2}[1,1,1]^T,\tfrac{1}{2}[1,-1,1]^T, \tag{100}$$
$$[1,0,0]^T,[0,0,0]^T\}.$$

Given the particular set of effects we started with, it follows that $\mathsf{Cone}[\mathcal{E}]$ is precisely the cone of effects corresponding to the Boxworld system.

Now we want to decide whether the statistics generated by every possible pair state-effect in $(\Omega,\mathcal{E})$ can admit a classical explanation; that is, we want to decide whether the statistics generated by the Boxworld system under any possible state preparation and effect can be explained by an underlying ontological model. So let us apply our linear program technique from Linear Program 1. On the one hand, the linear maps H_Ω and $H_{\mathcal{E}}$ take the form:

$$H_\Omega=\begin{bmatrix}1&1&1\\1&1&-1\\1&-1&1\\1&-1&-1\end{bmatrix},\quad H_{\mathcal{E}}=\begin{bmatrix}1&0&1\\1&0&-1\\1&1&0\\1&-1&0\end{bmatrix}. \tag{101}$$

On the other hand, the inclusion maps I_Ω and $I_{\mathcal{E}}$ satisfy $I_\Omega=I_{\mathcal{E}}=\mathbb{1}_3$.

Providing these as inputs in the linear program, one can find that there does not exist any σ with non-negative entries such that

$$\mathbb{1}_3=H_{\mathcal{E}}^T\cdot\sigma\cdot H_\Omega. \tag{102}$$

The maximally mixed state for a Boxworld system is given by $\mu=[1,0,0]^T$. With this specification, our linear program computes that the minimal amount r of completely depolarising noise which must be added to the experiment until it becomes simplex embeddable is $r=0.5$. That is, for the noisy scenario defined by the original set of effects, together with the original set of states but subjected to depolarizing noise[6] with probability $r=0.5$, one finds that Linear Program 1 is satisfiable, for

$$\sigma=\tfrac{1}{8}\begin{bmatrix}1&0&1&0\\0&1&0&1\\1&1&0&0\\0&0&1&1\end{bmatrix}. \tag{103}$$

[6] One could equally well take the original set of states and incorporate the noise into the effects instead.

The simplex embedding is found from σ (following Eq. (43) and Section C I) to be given by

$$\tau_{\Omega^{\mathcal{A}}} = \frac{1}{4}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & -1\\ 1 & -1 & 1\\ 1 & -1 & -1\end{bmatrix}, \quad \text{and} \quad \tau_{\mathcal{E}^{\mathcal{A}}} = \frac{1}{2}\begin{bmatrix}2 & 1 & 1\\ 2 & 1 & -1\\ 2 & -1 & 1\\ 2 & -1 & -1\end{bmatrix}. \tag{104}$$

From this, one can find an ontological model for the scenario (following Section C II). The model has four ontic states, over which the epistemic states for the four Boxworld states in Eq. (99) are

$$[\tfrac{1}{2},\tfrac{1}{2},0,0]^T, \quad [\tfrac{1}{2},0,\tfrac{1}{2},0]^T, \quad [0,0,\tfrac{1}{2},\tfrac{1}{2}]^T, \quad [0,\tfrac{1}{2},0,\tfrac{1}{2}]^T, \tag{105}$$

and the response functions for the six effects in Eq. (100) are

$$[0,\tfrac{1}{2},\tfrac{1}{2},1]^T, \quad [\tfrac{1}{2},1,0,\tfrac{1}{2}]^T, \quad [1,\tfrac{1}{2},\tfrac{1}{2},0]^T, \quad [\tfrac{1}{2},0,1,\tfrac{1}{2}]^T, \quad [1,1,1,1]^T, \quad [0,0,0,0]^T, \tag{106}$$

respectively. One can directly check that this model reproduces the noisy Boxworld statistics (for $r = 0.5$).

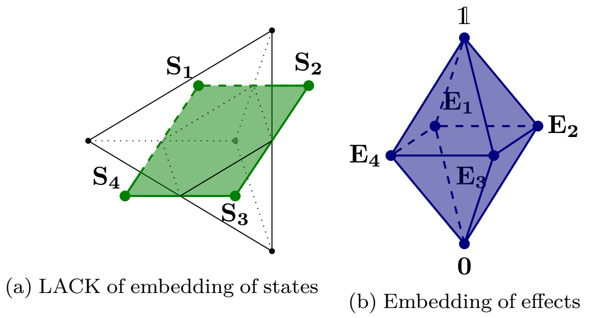


FIG. 1: LACK of a classical explanation for Example 4. (a) Depiction of the states in Ω (green dots), which fail to be embedded in a 3-dimensional slice of a 4-dimensional simplex. (b) Depiction of the effects in $\mathcal{E}$ (blue dots), embedded in a 3-dimensional slice of the 4-dimensional hypercube that is dual to the simplex in (a). Note that the convex hull of the effects happens to cover the entire hypercube in this particular slice. Notice that for this particular choice of embedding of effects, that the states must be represented (as shown in (a)) outside of the simplex, and so this does not constitute a simplex embedding. Because this is true for all possible embeddings of the effects, there is no possible noncontextual ontological model for—and hence classical explanation of—the operational scenario.

G. OPEN-SOURCE CODE

This article is accompanied by a Mathematica™ notebook which automates the assessment of simplex embeddability. The supplied notebook has been tested only for Mathematica versions 12 and 13, and it requires an installation of the `cdd` binary, available for Ubuntu linux (and therefore for Windows via WSL) at https://packages.ubuntu.com/jammy/amd64/libcdd-tools/filelist and for Mac at https://formulae.brew.sh/formula/cddlib.

I. How to use the code

The code's main user-accessible function is `DiscoverEmbedding`. This function can take inputs in two different formats, depending on whether one's starting point is a fragment of quantum theory or a GPT fragment more generally. That is, one may input

i. a set of density matrices and a set of POVM elements, or

ii. a set of GPT states, a set of GPT effects, a specification of the GPT unit effect, and a specification of the maximally mixed state in the GPT.

If one is only interested in answering the *qualitative* question of whether or not a given GPT fragment is nonclassical, then the specification of the maximally mixed state in the GPT is irrelevant. (The code is written to reflect this—if one inputs a GPT fragment but does not specify any maximally mixed state, it will still output the value of r, which determines whether the scenario is classically-explainable or not.)

The formatting of these inputs is as follows. The function takes four arguments in the case of a GPT fragment, but only the first two are needed for the case of a quantum fragment. The four arguments are: a set of states, a set of effects, a unit effect, and a maximally mixed state. For the case of a quantum fragment, the set of states is an order-3 tensor, i.e., a comma-separated list of density operators as $n \times n$ complex-valued matrices. Similarly, the set of effects is then a list of POVM elements, each being an $n \times n$ complex-valued matrix. No unit effect need be specified for the case of a quantum fragment, as it is automatically presumed to be the $n \times n$ identity matrix. Similarly, the maximally mixed state is implicit when the input is a quantum fragment.

For the case of a GPT fragment, one inputs a matrix $V_{\Omega^{\mathcal{F}}}$ for the set of states, a matrix $V_{\mathcal{E}^{\mathcal{F}}}$ for the set of effects, a row vector $u^{\mathcal{F}}$ for the unit effect, and a row vector $m^{\mathcal{F}}$ for the maximally mixed state. The four arguments to the function `DiscoverEmbedding` are separated by commas. $V_{\Omega^{\mathcal{F}}}$ should be constructed so that each of its rows is one of the GPT state vectors. Similarly, $V_{\mathcal{E}^{\mathcal{F}}}$ should be constructed so that each of its rows is one of the GPT effect vectors.

In both cases, the most important information returned by the code is the value of r. If $r = 0$, then it follows that the scenario in question is classically-explainable, and the code returns such a classical explanation in the form of an ontological model. Namely, it returns a specification the epistemic states and the response functions which represent the states and effects (respectively) in one's input. By necessity [5, 23], this ontological model constitutes a *noncontextual* ontological model for whatever operational scenario is described by the quantum or GPT input to the program. If $r > 0$, then one has witnessed nonclassicality for one's operational scenario, and the value of r is an operational measure of how nonclassical it is. The code again returns a noncontextual ontological model—not for the original scenario, but for the scenario after completely depolarizing noise is applied with probability r.

The output ontological model is given as two matrices. The rows of the first matrix μ are the epistemic states, where the ith row is a representation of the ith quantum (GPT) state. The rows of the second matrix ξ are the response functions, where the ith row is a representation of the ith quantum (GPT) effect.

The code also outputs a number of other possibly useful objects. All of these are described in the following sections.

II. Useful matrix formulations

We now recast the diagrammatic results proven above in a language which is more amenable to coding, and so is useful for seeing how our proofs connect with our open-source code. Along the way, we highlight three equivalent formulations of simplicial-cone embedding and formulate these in terms of matrix computations. Respectively, these formulations are statements of

1. The existence of a simplicial-cone embedding for a GPT fragment.
2. The existence of a simplicial-cone embedding for an accessible GPT fragment.
3. The satisfaction of our linear program.

We begin with a formulation that simply states the definition of simplicial-cone embeddability of a GPT fragment. Recall from the start of Section B III that a GPT fragment is simply a subset $\Omega^{\mathcal{F}}$ of states and a subset $\mathcal{E}^{\mathcal{F}}$ of effects from some underlying GPT. Recall also that a simplicial-cone embedding for a fragment is the same as that for the underlying GPT, Def. 1, but with the $\Omega^{\mathcal{G}}$ replaced by $\Omega^{\mathcal{F}}$ and with $\mathcal{E}^{\mathcal{G}}$ replaced by $\mathcal{E}^{\mathcal{F}}$.

As a first step, as discussed above, we encode all of the states in $\Omega^{\mathcal{F}}$ in a single matrix $V^T_{\Omega^{\mathcal{F}}}$, which one defines by picking a basis for S, finding the coordinates of each GPT state vector in that basis, and then taking the resulting set of coordinate-vectors as the columns of $V^T_{\Omega^{\mathcal{F}}}$. Analogously, define a single matrix $V_{\mathcal{E}^{\mathcal{F}}}$ which characterizes the set of effects $\mathcal{E}^{\mathcal{F}}$.

Then, one can formulate simplicial-cone embedding of a GPT fragment as

Formulation 1. [Compare to Definition 3]
Given a pair of matrices $V_{\Omega^{\mathcal{F}}}$ and $V_{\mathcal{E}^{\mathcal{F}}}$, does there exist a pair of matrices $\tau_{\Omega^{\mathcal{F}}}$ and $\tau_{\mathcal{E}^{\mathcal{F}}}$ such that

$$\tau_{\Omega^{\mathcal{F}}} \cdot V^T_{\Omega^{\mathcal{F}}} \geq_e 0, \tag{107a}$$
$$\tau_{\mathcal{E}^{\mathcal{F}}} \cdot V^T_{\mathcal{E}^{\mathcal{F}}} \geq_e 0, \tag{107b}$$
$$V^T_{\mathcal{E}^{\mathcal{F}}} \cdot V_{\Omega^{\mathcal{F}}} = V^T_{\mathcal{E}^{\mathcal{F}}} \cdot \tau^T_{\mathcal{E}^{\mathcal{F}}} \cdot \tau_{\Omega^{\mathcal{F}}} \cdot V_{\Omega^{\mathcal{F}}} \, ? \tag{107c}$$

As shown in Proposition 1, a GPT fragment is simplicial-cone embeddable if and only if the associated *accessible* GPT fragment is simplicial-cone embeddable.

To write a matrix formulation of the simplicial-cone embeddability of an accessible GPT fragment, we introduce matrices $V_{\Omega^{\mathcal{A}}}$ and $V_{\mathcal{E}^{\mathcal{A}}}$ which collect together vector representations of states and effects in the accessible GPT fragment into a single matrix as we did for the case of GPT fragments. The operational probabilities are then computed according to the bilinear map $B = I^T_{\mathcal{E}^{\mathcal{A}}} \cdot I_{\Omega^{\mathcal{A}}}$ introduced in Section B III.

We can then express simplicial-cone embeddability of accessible GPT fragments in terms of these matrices as

Formulation 2. [Compare to Definition 4]
Given a pair of matrices $V_{\Omega^{\mathcal{A}}}$ and $V_{\mathcal{E}^{\mathcal{A}}}$ with a probability map B, does there exist a pair of matrices $\tau_{\Omega^{\mathcal{A}}}$ and $\tau_{\mathcal{E}^{\mathcal{A}}}$ such that

$$\tau_{\Omega^{\mathcal{A}}} \cdot V^T_{\Omega^{\mathcal{A}}} \geq_e 0, \tag{108a}$$
$$\tau_{\mathcal{E}^{\mathcal{A}}} \cdot V^T_{\mathcal{E}^{\mathcal{A}}} \geq_e 0, \tag{108b}$$
$$B = \tau^T_{\mathcal{E}^{\mathcal{A}}} \cdot \tau_{\Omega^{\mathcal{A}}} \, ? \tag{108c}$$

Our third and final formulation explicitly constitutes a linear program. Its equivalence to the previous formulation follows from Proposition 1.

In terms of matrices, it is given by

Formulation 3. [Compare to Theorem 6]
Given a pair of matrices $V_{\Omega^{\mathcal{A}}}$ and $V_{\mathcal{E}^{\mathcal{A}}}$ which compose to form probabilities according to the bilinear map $B = I^T_{\mathcal{E}^{\mathcal{A}}} \cdot I_{\Omega^{\mathcal{A}}}$, does there exist any nonnegative matrix $\sigma \geq_e 0$ such that

$$B = H^T_{\mathcal{E}^{\mathcal{A}}} \cdot \sigma \cdot H_{\Omega^{\mathcal{A}}} \, ? \tag{109}$$

All the formulations above refer to the existence of *simplicial-cone* embeddings. We showed diagrammatically how to transform such an embedding into a *simplex embedding* in Appendix C I. We now express this transformation in the language of this section. A simplex embedding is more restrictive than a simplicial-cone embedding in that it additionally requires that $\tau_{\mathcal{E}^{\mathcal{F}}}$ is such that $\tau_{\mathcal{E}^{\mathcal{F}}} \cdot u_{\mathcal{F}}$ equals a vector (of length equal to the dimension of the simplex embedding) consisting of only ones. Given a *simplicial-cone* embedding, one can readily construct a simplex embedding by rescaling row i of the simplicial-cone embedding's matrix $\tau_{\Omega^{\mathcal{F}}}$ by $\tau_{\mathcal{E}^{\mathcal{F}}}[i]^T \cdot u_{\mathcal{F}}$, where $\tau_{\mathcal{E}^{\mathcal{F}}}[i]$ is the i-th column of $\tau_{\mathcal{E}^{\mathcal{F}}}$, and rescaling row i of the simplicial-cone embedding's $\tau_{\mathcal{E}^{\mathcal{F}}}$ by $\frac{1}{\tau_{\mathcal{E}^{\mathcal{F}}}[i] \cdot u_{\mathcal{F}}}$, whenever $\tau_{\mathcal{E}^{\mathcal{F}}}[i] \cdot u_{\mathcal{F}} > 0$, and by truncating the vectors to remove the elements in which $\tau_{\mathcal{E}^{\mathcal{F}}}[i] \cdot u_{\mathcal{F}} = 0$.

III. Internals of the code

The function DiscoverEmbedding proceeds in four stages, at each stage printing useful information about the embeddability or nonembeddability of the quantum or GPT fragment. Moreover, there is an optional zeroth stage which is only relevant if the user provides a quantum fragment rather than a GPT fragment.

Stage 0 (when needed): Constructing a GPT fragment from a quantum fragment

If the input to the program is a specification of a set of quantum density matrices and a set of POVM elements, then there is a preliminary stage to the algorithm whereby the quantum fragment is recast in the GPT formalism.

First, we use the generalised Gell-Mann matrices [30] as a basis for the space of Hermitian operators to represent the density matrices and POVM elements as real vectors. In particular, note that we choose the scaling of the Gell-Mann matrices so that the trace inner-product for Hermitian operators is mapped to the dot-product for the real vectors. Second, we compute the unit effect and maximally mixed state by converting the Hermitian operators $\mathbb{1}$ and $\mathbb{1}/d$ into real vectors, again via the Gell-Mann matrices.

This provides all of the data necessary for the second kind of input to the algorithm, and so the algorithm is then called on the input in this form. The rest of the stages are therefore the same regardless of which kind of input is given.

Stage 1: Constructing an accessible GPT fragment from a GPT Fragment

The first task of our algorithm is to move from a GPT fragment to an *accessible* GPT fragment; see Eq. 21. Thus, we initially construct (and print) the inclusion matrices $I_{\mathcal{E}^{\mathcal{A}}}$ and $I_{\Omega^{\mathcal{A}}}$, as well as the projection matrices $P_{\mathcal{E}^{\mathcal{A}}}$ and $P_{\Omega^{\mathcal{A}}}$ given by their respective Moore-Penrose pseudoinverses. The inclusion matrices are critical for constructing the left-hand side of condition (11b) of the main text in our linear program. The projection matrices are useful for converting $\tau_{\Omega^{\mathcal{A}}}$ and $\tau_{\mathcal{E}^{\mathcal{A}}}$ to $\tau_{\Omega^{\mathcal{F}}}$ and $\tau_{\mathcal{E}^{\mathcal{F}}}$ respectively; see Eq. (27). Additionally, the projection matrices are used internally to construct an accessible GPT fragment, namely the three objects $V_{\Omega^{\mathcal{A}}}$, $V_{\mathcal{E}^{\mathcal{A}}}$, $u^{\mathcal{A}}$, which are then printed for the user.

Internally, the inclusion and projection matrices are formed using Mathematica's `RowReduce` and `Pseudoinverse` commands.

Stage 2: Computing the facets of the accessible GPT state and effect cones

Our linear program for characterizing the simplicial-cone embeddability of an accessible GPT fragment requires that we construct $H_{\Omega^{\mathcal{A}}}$ and $H_{\mathcal{E}^{\mathcal{A}}}$, e.g., for the right-hand side of Eq. (9b) of the main text. In particular, we define $H_{\Omega^{\mathcal{A}}}$ and $H_{\mathcal{E}^{\mathcal{A}}}$ as lists of the facet-defining inequalities, see Eqs. (2) and (6) of the main text or Eqs. (28) and (31).

The hypercone facet inequalities are derived internally by making an external call to the `cdd` binary. Note that cdd relies on the double description method [31] while in [2] they propose the use of the reverse search algorithm [32]. There are pros and cons to each of these algorithms, so in a future version we intend to allow the user to choose which to use.

The code then prints $H_{\Omega^{\mathcal{A}}}$ and $H_{\mathcal{E}^{\mathcal{A}}}$.

Stage 3: Finding the robustness of nonclassicality

Next, the code implements the linear program described in Linear Program 2 in the main text. We explicitly return the minimum value of the primal objective, namely r, as well as a corresponding nonnegative matrix σ. To do so the maximally-mixed state is used to construct the depolarization map $D^{\mathcal{F}} := m^{\mathcal{F}^T} \cdot u^{\mathcal{F}}$. When no maximally-mixed state is explicitly provided by the user—or, in the case where the input is quantum, computed, as in stage 0—the code will automatically generate a "central" state by taking the uniform mixture of all states in $V_{\Omega^{\mathcal{F}}}$. Recall that the particular noise model implemented by our code depends on the choice of $m^{\mathcal{F}}$, and so the precise value of r is only physically meaningful relative to this noise model; see again the discussion in Section D. For any choice, however, the input GPT fragment is classical simplex-cone embeddable if and only if $r = 0$.

If our program finds that the smallest r is strictly positive, the code will further print out the matrix formed by the left-hand side (equivalently, by the right-hand side) of Eq. (11b) of the main text. (This information may be of interest insofar as it encodes the effective probability rule for the accessible GPT fragment generated after the depolarization.)

Stage 4: Conversion to a simplex embedding from a simplicial-cone embedding

Finally, the algorithm returns matrices for both $\tau_{\Omega^{\mathcal{A}}}$ and $\tau_{\mathcal{E}^{\mathcal{A}}}$, rescaled so that $\tau_{\mathcal{E}^{\mathcal{A}}} \cdot u^{\mathcal{A}} = [1,1,...,1]$. That is, the final printout from the code is an explicit simplex embedding of the (depolarized, if necessary) accessible GPT fragment, per Eq. (25).

Note that if $r > 0$, the final simplex embedding given by $\tau_{\Omega^{\mathcal{A}}}$ and $\tau_{\mathcal{E}^{\mathcal{A}}}$ is actually not a simplex embedding of the *original* accessible GPT fragment, but rather it is a simplex embedding of the *minimally depolarized* accessible GPT fragment which *is* simplicial-cone embeddable. Recall that per Proposition 1 and Appendix C I, an accessible GPT fragment is simplex embeddable whenever it is simplicial-cone embeddable, so there is not loss of generality in returning simplex embeddings as opposed to simplicial-cone embeddings.

Finally, an explicit ontological model for one's scenario is constructed, following the construction in Section C II. This model is specified as a matrix μ whose rows form the epistemic states representing the states in (the rows of) $V_{\Omega^{\mathcal{F}}}$, together with a matrix ξ whose rows form the response functions representing the effects in (the rows of) $V_{\mathcal{E}^{\mathcal{F}}}$.

[1] F. Shahandeh, M. J. W. Hall, and T. C. Ralph, "Measurement-Device-Independent Approach to Entanglement Measures," Phys. Rev. Lett. **118**, 150505 (2017).

[2] V. Gitton and M. P. Woods, "Solvable criterion for the contextuality of any prepare-and-measure scenario," Quantum **6**, 732 (2022).
[3] M. P. Müller and A. J. P. Garner, "Testing quantum theory by generalizing noncontextuality," Phys. Rev. X **13**, 041001 (2023).
[4] F. Shahandeh, "Contextuality of general probabilistic theories," PRX Quantum **2**, 010330 (2021).
[5] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, "Characterization of noncontextuality in the framework of generalized probabilistic theories," PRX Quantum **2**, 010331 (2021).
[6] R. W. Spekkens, "Contextuality for preparations, transformations, and unsharp measurements," Phys. Rev. A **71**, 052108 (2005).
[7] M. F. Pusey, L. Del Rio, and B. Meyer, "Contextuality without access to a tomographically complete set," arXiv:1904.08699 (2019).
[8] M. D. Mazurek, M. F. Pusey, K. J. Resch, and R. W. Spekkens, "Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories," PRX Quantum **2**, 020302 (2021).
[9] M. J. Grabowecky, C. A. J. Pollack, A. R. Cameron, R. W. Spekkens, and K. J. Resch, "Experimentally bounding deviations from quantum theory for a photonic three-level system using theory-agnostic tomography," Phys. Rev. A **105**, 032204 (2022).
[10] R. W. Spekkens, "Evidence for the epistemic view of quantum states: A toy theory," Phys. Rev. A **75**, 032110 (2007).
[11] R. W. Spekkens, "Quasi-Quantization: Classical Statistical Theories with an Epistemic Restriction," in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer Netherlands, Dordrecht, 2016) pp. 83–135.
[12] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, "Reconstruction of Gaussian quantum mechanics from Liouville mechanics with an epistemic restriction," Phys. Rev. A **86**, 012103 (2012).
[13] L. Catani, M. Leifer, D. Schmid, and R. W. Spekkens, "Why interference phenomena do not capture the essence of quantum theory," Quantum **7**, 1119 (2023).
[14] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, "Accessible fragments of generalized probabilistic theories, cone equivalence, and applications to witnessing nonclassicality," Phys. Rev. A **107**, 062203 (2023).
[15] M. Müller, "Probabilistic theories and reconstructions of quantum theory," SciPost Physics Lecture Notes , 028 (2021).
[16] L. Lami, "Non-classical correlations in quantum mechanics and beyond," arXiv preprint arXiv:1803.02902 (2018).
[17] M. Plávala, "General probabilistic theories: An introduction, Physics Reports **1033**,1-64 " (2023).
[18] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Quantum from Principles," in *Fundamental Theories of Physics* (Springer Netherlands, 2015) pp. 171–221.
[19] L. Hardy, "Reformulating and reconstructing quantum theory," arXiv:1104.2066 (2011).
[20] L. Hardy, "Quantum Theory From Five Reasonable Axioms," arXiv:quant-ph/0101012 (2001).
[21] J. Barrett, "Information processing in generalized probabilistic theories," Phys. Rev. A **75**, 032304 (2007).
[22] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Probabilistic theories with purification," Phys. Rev. A **81**, 062348 (2010).
[23] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, "A structure theorem for generalized-noncontextual ontological models," (2020), arXiv:2005.07161 [quant-ph].
[24] B. Coecke, T. Fritz, and R. W. Spekkens, "A mathematical theory of resources," Info. Comp. **250**, 59 (2016).
[25] T. Gonda and R. W. Spekkens, "Monotones in General Resource Theories," (2023).
[26] R. Kunjwal and R. W. Spekkens, "From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism," Phys. Rev. Lett. **115**, 110403 (2015).
[27] I. Marvian, "Inaccessible information in probabilistic models of quantum systems, non-contextuality inequalities and noise thresholds for contextuality," (2020), arXiv:2003.05984 [quant-ph].
[28] V. P. Rossi, D. Schmid, J. H. Selby, and A. B. Sainz, "Contextuality with vanishing coherence and maximal robustness to dephasing," Phys. Rev. A **108**, 032213 (2023).
[29] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Nonlocal correlations as an information-theoretic resource," Physical Review A **71**, 022101 (2005).
[30] G. Kimura, "The bloch vector for n-level systems," Physics Letters A **314**, 339 (2003).
[31] T. S. Motzkin, H. Raiffa, G. L. Thompson, and R. M. Thrall, "The double description method," Contributions to the Theory of Games **2**, 51 (1953).
[32] D. Avis and K. Fukuda, "A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra," in *Proceedings of the seventh annual symposium on Computational geometry* (1991) pp. 98–104.

Contextuality with vanishing coherence and maximal robustness to dephasing

Vinicius P. Rossi,* David Schmid, John H. Selby, and Ana Belén Sainz
International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-309 Gdańsk, Poland

(Received 17 January 2023; accepted 25 August 2023; published 18 September 2023)

Generalized contextuality is a resource for a wide range of communication and information processing protocols. However, contextuality is not possible without coherence, and so can be destroyed by dephasing noise. Here, we explore the robustness of contextuality to partially dephasing noise in a scenario related to state discrimination (for which contextuality is a resource). We find that a vanishing amount of coherence is sufficient to demonstrate the failure of noncontextuality in this scenario, and we give a proof of contextuality that is robust to arbitrary amounts of partially dephasing noise. This is in stark contrast to partially depolarizing noise, which is always sufficient to destroy contextuality.

DOI: 10.1103/PhysRevA.108.032213

I. INTRODUCTION

Understanding what is nonclassical about quantum theory is crucial for determining which tasks can be optimally performed with quantum resources. One quantum resource that is useful in many tasks within computation [1], communication [2], information processing [3–6], metrology [7], cloning [8], and state discrimination [9–12], is generalized contextuality [13] (henceforth referred to simply as "contextuality").

A given experiment is said to be a proof of contextuality when its statistics are incompatible with the existence of a noncontextual ontological model, i.e., models wherein one's ontology is a set (of classical states), dynamics are represented as functions, where inferences are made using Bayesian probability theory and Boolean logic, and where a methodological version of the assumption of *Leibnizianity* [14] is satisfied. This assumption stipulates that the explanation for procedures being indiscernible at the operational level is that they are also indiscernible at the ontological level [15,16]. This notion of nonclassicality—the nonexistence of a noncontextual ontological model—was proven to be equivalent to other notions of nonclassicality, such as the nonexistence of a quasiprobability representation in quantum optics [17,18], and the nonexistence of a simplex embedding in generalized probabilistic theories [19]. Furthermore, this notion of nonclassicality is closely related to the notions arising in the study of quantum Darwinism [20], macrorealism [21], Bell scenarios [22–24], and the detection of anomalous weak values [25]. In our view, generalized contextuality is our most well-motivated notion of nonclassicality.

Of particular interest to this work is the aforementioned notion of simplex embeddability. This simple geometric characterization of the notion of noncontextuality within the framework of generalized probabilistic theories (GPTs) has been useful for exploring the relationship between contextuality and incompatibility [26,27]. It has also been employed in the development of an open-source code for testing whether a given prepare-and-measure scenario constitutes a proof of contextuality, and, moreover, for providing a quantification of how robust to depolarizing noise this proof is [28]. We will leverage this tool here in our study of the relationship between contextuality and coherence.

It is well known that contextuality is always destroyed by partial (but sufficiently large) depolarizing noise.[1] However, the question of how robust contextuality is to dephasing noise has not previously been studied. Note that the existence of coherence does not immediately imply contextuality, since it is present in epistemically restricted theories [16,31–33] for which noncontextual ontological models are known to exist. On the other hand, contextuality is not possible without coherence, a fact that we prove explicitly in Appendix B. However, this leaves open the question of how contextuality is affected by *partial* dephasing noise. This question is of particular importance given that decoherence theory [34,35] shows that dephasing noise arises in generic open system dynamics.

In this work, we show that there are proofs of contextuality that can be obtained with *any* nonzero amount of coherence. We then modify the open-source linear program from Ref. [28] and use this to investigate the robustness of contextuality to the action of dephasing noise with respect to a fixed basis in a collection of prepare-and-measure scenarios. Finally, we find a proof of contextuality that is maximally robust to dephasing noise, in the sense that the experiment remains a proof of contextuality for any amount of decoherence apart from total decoherence.

II. PRELIMINARY NOTIONS

The broad range of experiments we are interested in investigating consists of those that can be thought of as preparing

[1]This fact has been noted in particular scenarios [9,29,30], and in general scenarios it follows immediately from simplex embeddability.

*prettirossi.vinicius@gmail.com

a system in a laboratory in a variety of different ways and probing it with a variety of measurements. Such experiments are known as *prepare-and-measure scenarios*. These can be studied from a theory-agnostic viewpoint, where only a minimal set of *operational* elements (i.e., properties or objects that are manifestly observable) are used to describe it. In such situations, one can analyze the experimental scenario without making any assumptions about the nature of the system in question, e.g., what its intrinsic properties are or how it behaves. Rather, one simply focuses on (i) the classical labels of the ways in which one may prepare this system ($P \in \mathcal{P}$), (ii) the classical labels of the measurement procedures one may perform ($M \in \mathcal{M}$), (iii) the classical labels of the outcomes $k \in K$ of the measurement procedures, (iv) the resulting statistics $\{p(k|M,P)\}_{[k|M]\in K\times\mathcal{M},P\in\mathcal{P}}$ of the experiment, and (v) the operational equivalences between different preparation or measurement outcome procedures, denoted $\simeq$ and defined in the next paragraph. The tuple $(\mathcal{P},\mathcal{M},K,p,\simeq)$ that captures the information analyzed in such a prepare-and-measure scenario is often referred to as an *operational scenario*.

An *operational theory* is the set of all possible realizable preparations, measurements, and outcomes and their respective statistics, i.e., it is the maximal operational scenario for a given system. Often a given operational theory allows for a kind of equivalence between different procedures: sometimes two preparation procedures P and P' yield the same statistics for any possible measurement outcomes, or two measurement outcomes $[k|M]$ and $[k'|M']$ do so for any possible preparation. When this happens, we say that P and P' (respectively, $[k|M]$ and $[k'|M']$) are *inferentially equivalent*, denoted by $P \simeq P'$ (respectively, $[k|M] \simeq [k'|M']$). Here, it is crucial to assess such equivalences with respect to the full operational theory, and not only relative to the specific preparations or measurement outcomes in the specific operational scenario under investigation. That is, we define the operational equivalence relation for an operational scenario as the one that it inherits from the operational theory in which it lives.

Since inferentially equivalent processes cannot be distinguished by the operational predictions p they generate, it is often useful to discard this *context* information by identifying inferentially equivalent processes with a single representative of the group. This operation is termed *quotienting* [36] and provides the way to construct a so-called GPT [37–39] for a corresponding operational theory. See Appendix C for a concrete definition of a GPT.

An *ontological model* seeks to explain the observed statistics in one's scenario by associating (i) the system to some set of ontic states Λ, (ii) preparations P to *epistemic states*, μ_P, which are probability distributions over Λ, and (iii) measurement-outcome pairs $[k|M]$ in the operational theory to *response functions* $\xi_{k|M}$ over Λ, such that $p(k|M,P) = \sum_{\lambda\in\Lambda}\xi_{k|M}(\lambda)\mu_P(\lambda)$. An ontological model is said to be noncontextual if inferentially equivalent preparations are mapped to the same epistemic state, and inferentially equivalent measurement-outcome pairs are mapped to the same response function.[2] For the purpose of this paper, it suffices to know that the notion of a noncontextual ontological model for the operational theory has an equivalent characterization at the level of the GPT related to it via quotienting. That is, a GPT is associated to an operational theory that is noncontextual if and only if the GPT is simplex embeddable [19]. Intuitively, such GPTs have a state space that fits inside a simplex, and an effect space that fits inside the dual to the simplex. We define simplex embeddability formally in Appendix C.

The existence of a simplex embedding can be tested using the linear program introduced in Ref. [28]. In the case of quantum theory (which is the case we study here) the linear program simply takes as input a set of density matrices (representing the preparations) and a set of positive operator-valued measurement (POVM) elements (representing the measurement-outcome pairs), and checks whether or not these are simplex embeddable, and consequently, whether the quantum scenario admits a noncontextual ontological model. Furthermore, in the case that the code fails to find a simplex embedding, it computes how much depolarizing noise must be added to the input states (or equivalently, measurements) such that a simplex embedding is found. In this work, we are interested in studying quantum prepare-and-measure scenarios under the action of *dephasing* rather than depolarizing noise, so we modify the code from Ref. [28] to estimate the robustness to dephasing rather than depolarization. A summary of how the code works and of our modifications to it is given in Appendix C.

III. PROOF OF CONTEXTUALITY WITH VANISHING COHERENCE

Inspired by Ref. [9], which demonstrates that contextuality is a resource powering an advantage for minimum-error state discrimination (MESD), we focus on a prepare-and-measure scenario constructed from the MESD scenario. Our scenarios of interest consist of four preparations $\{P_\psi, P_{\bar{\psi}}, P_\phi, P_{\bar{\phi}}\}$ and three binary measurements $\{M_\psi, M_\phi, M_g\}$. The preparations consist of pure states $|\psi\rangle$, $|\phi\rangle$ of a qubit system, and their orthogonal counterparts. That is, $P_\psi \to |\psi\rangle\langle\psi|$ and $P_{\bar{\psi}} \to |\bar{\psi}\rangle\langle\bar{\psi}|$, with $\langle\psi|\bar{\psi}\rangle = 0$ (and similarly for P_ϕ). Measurements M_ψ and M_ϕ are simply projections onto $|\psi\rangle$ and $|\phi\rangle$, respectively, while M_g is the Helstrom measurement, comprised of projectors onto the basis that straddles $|\psi\rangle$ and $|\phi\rangle$ [40]. As all these preparations and measurements lie within a two-dimensional slice of the Bloch sphere, we can, without loss of generality, take this slice to be the ZX plane of the Bloch sphere. Furthermore, we fix our system of coordinates such that the projectors E_{g_ψ} and E_{g_ϕ} associated with the measurement M_g lie aligned to the Z axis. The preparations and measurements in the scenario can be parametrized by the angle $\theta \in [0, \frac{\pi}{2}]$ between any of the preparations and the Z axis, as shown in Fig. 1. We will consider dephasing relative to the Z axis.

Notice that the parameter θ is closely related to the amount of coherence (relative to the Z axis) in the preparation and

[2] It is for this reason that inferential equivalences should be assessed relative to the entire scope of possible procedures in the theory rather than those of the scenario; it does not make sense to impose a constraint on an ontological description that is contingent on what we happen to have chosen to do in a given experiment.

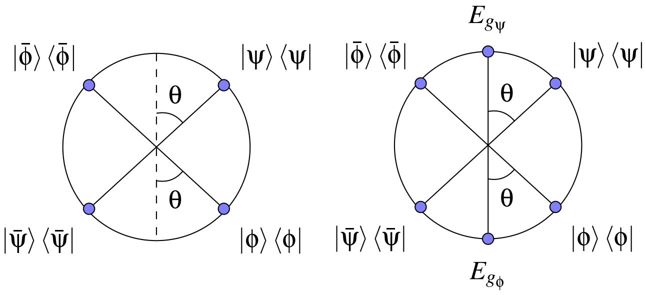


FIG. 1. Preparation (left) and measurement (right) procedures in the studied scenario, represented on a two-dimensional slice of the Bloch sphere. The vertical axis is taken to be the Z axis.

measurement procedures. If the coherence quantifier, C, is the trace distance of the state from the fully dephased version of the state [41], for instance, then we find that

$$C(|\psi\rangle\langle\psi|) := \left\| |\psi\rangle\langle\psi| - \sum_{i=0}^{1} |\langle i|\psi\rangle|^2 |i\rangle\langle i| \right\|_1 = \sin\theta, \quad (1)$$

with the same result for all preparations and measurements (other than E_{g_ψ} and E_{g_ϕ} for which the coherence is zero). Hence, increasing θ means increasing the coherence in both the states and effects.

Reference [9] discusses the consequences of noncontextuality for this scenario. Reference [9] shows that quantum theory allows for a higher probability of success at distinguishing $|\psi\rangle$ from $|\phi\rangle$ than is possible in any noncontextual theory. Hence, there is a quantum advantage for this task coming from contextuality. Reference [9] also analytically estimates how much depolarizing noise, $r_{\min}$, must be added to the quantum model until this quantum advantage disappears, i.e., until one's quantum measurements perform no better than noncontextual measurements could. Starting from the expression for depolarized effects

$$E \mapsto \mathcal{D}_r^{\text{depol}}(E) := (1-r)E + \frac{r}{2}\mathbb{1}, \quad (2)$$

imposing the existence of a noncontextual model for the quantum scenario implies that

$$r_{\min}^{\text{depol}} = 1 - \frac{1}{\sin^2\theta + \cos\theta}. \quad (3)$$

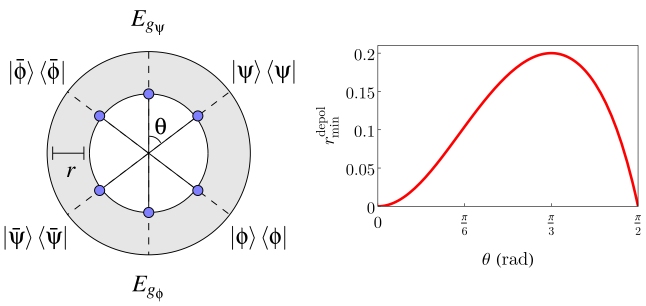


FIG. 2. Left: representation of the action of *depolarizing noise* on the scenario. Right: analytical plot of contextual robustness to depolarization as a function of the angle between the prepared states and the Z axis.

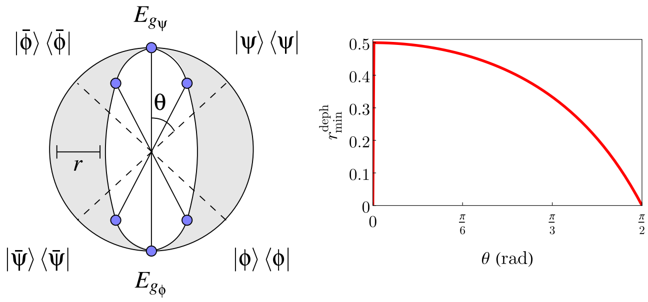


FIG. 3. Left: representation of the action of *dephasing noise* on the plane in which measurements live. Right: analytical plot of contextual robustness to dephasing as a function of the angle between the prepared states and the Z axis.

This was first computed in Ref. [9], although with a minor error that we correct in our proof in Appendix A.

This is plotted in Fig. 2, from which we can see that even small values of θ allow for a proof of contextuality. Indeed, the robustness to depolarization is null only for $\theta = 0$ and $\theta = \frac{\pi}{2}$, circumstances in which $|\psi\rangle$ is equal to either $|\phi\rangle$ or $|\bar{\phi}\rangle$. In this case, the cardinality of the sets of states/effects decreases, and simplex embeddability becomes possible. Since coherence vanishes as θ goes to zero, this establishes the following result (which was not previously recognized, although it requires only the results of Ref. [9] reiterated just above).

Result 1. There are proofs of the failure of noncontextuality that can be achieved in a prepare-and-measure scenario with vanishing (but nonzero) coherence among both the states and the effects.

This particular scenario is not very robust to depolarization noise. Even at its peak,[3] for $\theta = \frac{\pi}{3}$, the robustness to depolarization is 0.2. Moreover, the robustness goes smoothly to zero as the coherence goes to 0, and consequently is very small for small coherence. As we will see, this same scenario is considerably more robust to dephasing, and a closely related scenario is in fact maximally robust to dephasing.

IV. PROOF OF CONTEXTUALITY MAXIMALLY ROBUST TO DEPHASING

Next, we explore how contextuality behaves in this scenario under dephasing (rather than depolarizing) noise. In this case, the noisy projectors are given by

$$E \mapsto \mathcal{D}_r^{\text{deph}}(E) = (1-r)E + r\sum_{i\in\{0,1\}} \langle i|E|i\rangle|i\rangle\langle i|, \quad (4)$$

where $\{|i\rangle\}_{i\in\{0,1\}}$ is the Z basis. Imposing the existence of a noncontextual model for the scenario implies that

$$r_{\min}^{\text{deph}} = 1 - \frac{1-\cos\theta}{\sin^2\theta} \quad (5)$$

as proven in Appendix A and plotted in Fig. 3.

[3] Interestingly, this peak occurs at $\theta = \frac{\pi}{3}$, where one finds the same set of effects considered in Ref. [13] to obtain the first proof of the failure of measurement noncontextuality for POVMs in a two-dimensional system.

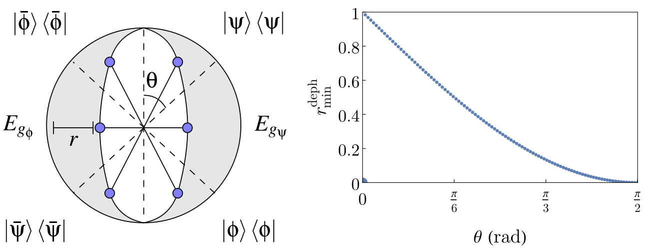


FIG. 4. Left: representation of the action of *dephasing noise* on the plane in which measurements of the rotated scenario live. Right: numeric estimation of contextual robustness to dephasing as a function of the angle between the prepared states and the Z axis, with the extra measurement now lying along the X axis.

Figure 3 shows that the amount of contextuality *decreases* monotonically as coherence increases (at least according to these measures of contextuality and coherence). While this might at first seem counterintuitive, one can understand it by noting that even under large dephasing noise, states and effects with little coherence to begin with are barely affected; that is, the dephasing channel is close to identity on such states and effects. However, this intuition only goes so far, as we will give an example below where one achieves twice the robustness to dephasing by including effects that have maximal coherence in the dephased basis. Just as was the case for depolarizing noise, the robustness to dephasing drops to zero when $|\psi\rangle = |\phi\rangle$, which implies a discontinuity in the plot at $\theta = 0$, where $r_{\min}^{\text{deph}}$ falls from 0.5 to 0.

Moreover, note that the maximum robustness relative to dephasing (0.5) is much higher than the maximal robustness to depolarization (0.2). A natural question is whether one can find scenarios where the contextual robustness to dephasing approaches its logical maximum, or whether (as for the contextual robustness to depolarization) this quantity is always bounded from above. In the following, we identify such a scenario by carrying out a numerical exploration using a modified version of the linear program from Ref. [28].

Consider modifying the above scenario by rotating the measurement M_g from the Z to the X axis, as in Fig. 4, thus maximizing the coherence in the effects associated with this measurement. In the case of depolarizing noise, this scenario is equivalent by symmetry to the original scenario under relabeling. Here, however, we are interested in the case of dephasing noise in the Z basis. (Note that this scenario is equivalent to the original scenario under relabeling and considering the noise to be in the X basis rather than the Z basis.) The robustness to dephasing for this scenario as a function of θ is plotted in Fig. 4. The most striking feature of this plot is that the robustness approaches 1 as $\theta \to 0$, so that the scenario achieves the maximum logically possible robustness to dephasing. This is in stark contrast with the original scenario (where the maximal dephasing robustness was 0.5).

Moreover, notice that if we start with the scenario described in Fig. 4 and then dephase by some r such that $1 - r$ is vanishingly small, then we can view the dephased scenario as a new scenario that has only a vanishing amount of coherence in the measurements, but which is still contextual and indeed is itself robust to arbitrary amounts of dephasing noise. (This follows from the fact that dephasing it by some r' is the same as dephasing the original scenario by $1 - (1 - r')(1 - r) > 0$.) Thus, we have established the following.

Result 2. There are proofs of the failure of noncontextuality that can be achieved in a prepare-and-measure scenario in the presence of arbitrarily large dephasing noise. One may moreover find some scenarios of this sort where the states and effects have vanishing (but nonzero) coherence.

Based on the intuitive arguments above, one may have expected the original scenario (Fig. 3) to be more robust than this rotated one (Fig. 4), since the highly coherent effects in the latter case are strongly affected by dephasing noise. However, this effect is clearly more than compensated for by the fact that the effects E_{g_ϕ} and E_{g_ψ}, in this case, are further from the other states and effects in the scenario, thus making it more difficult to find a simplex embedding of the GPT.

In Appendix D we present a numerical plot scanning between the case in which measurement M_g is aligned with the Z axis and when it is aligned with the X axis, showing that the latter is indeed the only scenario in this family for which maximal robustness is achieved.

A final natural question is how contextuality is affected in the presence of both dephasing and depolarizing noise. In Appendix D, we study contextual robustness to dephasing in a scenario where a small amount of depolarizing noise is added to the measurement M_g, and show that even when M_g is partially depolarized, proofs of contextuality can be obtained in the dephased scenario as long as the depolarizing noise on M_g does not surpass a bound given by the amount of coherence available from the states.

V. DISCUSSION

We have exhibited scenarios in which any nonzero amount of coherence is enough to prove the failure of the assumption of noncontextuality, and introduced an example in which a proof of contextuality can be robust to any amount of dephasing noise other than complete dephasing. This work showcases the versatility of the linear program from Ref. [28] and invites further research on how robust specific proofs of contextuality are under different types of noise.

Another recent work explored the connection between coherence and contextuality, using event graphs [42]. Violations of graph inequalities that witness both basis-independent coherence and contextuality [43,44] were derived in Ref. [42] and applied to a similar scenario [45], in this case with six preparations rather than four. These inequalities are not violated for the whole interval $\theta \in (0, \frac{\pi}{2})$, in contrast to the scenario studied herein. To get some intuition on why, notice that the existence of basis-independent coherence is not sufficient to guarantee the failure of noncontextuality. To see this, recall that contextuality always goes to zero by partial depolarization, and yet the only depolarizing process that destroys all basis-independent coherence is the totally depolarizing process.

ACKNOWLEDGMENTS

We thank Elie Wolfe for the ready support with the linear program functioning and Rafael Wagner for the

TABLE I. Possible statistics for an operational theory with four preparations and three binary measurements, satisfying the equivalence $\frac{1}{2}P_\psi + \frac{1}{2}P_{\bar{\psi}} = \frac{1}{2}P_\phi + \frac{1}{2}P_{\bar{\phi}}$. Note that we only show the $k = 0$ outcome statistic for each measurement as the $k = 1$ are then uniquely determined by normalization.

$p(k\|M, P)$	P_ϕ	P_ψ	$P_{\bar{\phi}}$	$P_{\bar{\psi}}$
$0\|M_\phi$	$1-\epsilon$	c	ϵ	$1-c$
$0\|M_\psi$	c	$1-\epsilon$	$1-c$	ϵ
$0\|M_g$	s	$1-s$	$1-s$	s

fruitful discussion. D.S. and V.R. were supported by the Foundation for Polish Science (IRAP project, ICTQT, Contract No. MAB/2018/5, cofinanced by EU within Smart Growth Operational Programme). J.H.S. was supported by the National Science Centre, Poland (Opus project, Categorical Foundations of the Non-Classicality of Nature, Project No. 2021/41/B/ST2/03149). A.B.S. acknowledges support by the Digital Horizon Europe project FoQaCiA, Foundations of quantum computational advantage, GA No. 101070558, funded by the European Union, NSERC (Canada), and UKRI (UK). Some figures were prepared using TIKZIT.

APPENDIX A: ANALYTICAL DERIVATION OF THE ROBUSTNESS TO DEPOLARIZATION AND DEPHASING

1. Robustness to depolarization

The prepare-and-measure scenario of interest for our purposes consists of four preparations denoted $P \in \{P_\psi, P_{\bar{\psi}}, P_\phi, P_{\bar{\phi}}\} =: \mathcal{P}$ and three binary-outcome measurements $k|M \in \{0, 1\} \times \{M_\phi, M_\psi, M_g\} =: K \times \mathcal{M}$. Without any further restrictions, there would be 12 free parameters in the data table $\{p(k|M, P)\}_{P\in\mathcal{P},[k|M]\in K\times\mathcal{M}}$. However, Ref. [9] imposes further constraints on these preparations and measurement outcomes, based on the symmetries of $\mathcal{P}$ and $\mathcal{M}$ and on the inferential equivalence

$$\tfrac{1}{2}P_\psi + \tfrac{1}{2}P_{\bar{\psi}} = \tfrac{1}{2}P_\phi + \tfrac{1}{2}P_{\bar{\phi}}. \tag{A1}$$

As a result, any data table satisfying these will necessarily have only three free parameters. In particular, Ref. [9] takes the three parameters to be denoted by s, c, and ϵ, which are related to the statistics of the operational theory as represented in Table I.

The demand that there is a noncontextual ontological model for this operational theory forces a relation upon these three parameters. The derivation of this relation is well explained in Ref. [9] and culminates in solving a rather extensive linear system, and is given by the inequality

$$s \leqslant 1 - \frac{c-\epsilon}{2}. \tag{A2}$$

If a data table parametrized as in Table I satisfies this condition, then it can be explained by a generalized noncontextual ontological model, otherwise it cannot.

In our quantum experiment, described in Fig. 1 of the main text, we find that the inferential equivalence (A1) is satisfied and that we can write the three parameters in the data table as functions of θ and r. To do so, first note that the depolarizing noise acts on the effects of the scenario such that

$$E \mapsto \mathcal{D}_r^{\text{depol}}(E) := (1-r)E + \frac{r}{2}\mathbb{1}, \tag{A3}$$

where $\mathbb{1}$ is the 2×2 identity matrix. This allows us to compute the statistics of our experiment for depolarized measurements in Table II.

By comparing the two tables we can write s, c, and ϵ as functions of θ and r:

$$s = \frac{1}{2} + \frac{1-r}{2}\cos\theta; \tag{A4}$$

$$c = (1-r)\sin^2\theta + \frac{r}{2}; \tag{A5}$$

$$\epsilon = \frac{r}{2}. \tag{A6}$$

From this, we can compute the minimal amount of depolarizing noise $r_{\min}^{\text{depol}}$, which is required so that the quantum experiment admits of a noncontextual ontological model. That is, the value of r such that the inequality (A2) is satisfied tightly. The value obtained for $r_{\min}^{\text{depol}}$ will depend on θ and so we obtain the equation

$$r_{\min}^{\text{depol}} = 1 - \frac{1}{\sin^2\theta + \cos\theta}, \tag{A7}$$

which is the equation that was shown and discussed in the main text. Notice that in Ref. [9] the depolarizing noise acts both on states and on effects, so the statistics in Table II in our case provide different values for the parameters s, c, and ϵ than in that work. However, the plots of $r_{\min}^{\text{depol}}$ coincide here and in Ref. [9] (up to a reparametrization of θ) due to a miscalculation in the latter, which rescaled $r_{\min}^{\text{depol}}$ to the case of depolarizing noise acting only on the effects. Because the two cases—depolarizing noise acting only on effects or acting on both effects and measurements—are equivalent up to this reparametrization, there is no impact on any of the analyses in Ref. [9].

2. Robustness to dephasing

We can now repeat the same analysis from the previous section, but in this case considering the scenario described in Fig. 3 of the main text, where r parametrizes the amount of dephasing rather than depolarizing noise. Recall that in this scenario the dephased effects are given by

$$E \mapsto \mathcal{D}_r^{\text{deph}}(E) = (1-r)E + r\sum_{i\in\{0,1\}}\langle i|E|i\rangle|i\rangle\langle i|, \tag{A8}$$

where $\{|i\rangle\}_{i\in\{0,1\}}$ is the Z basis. The statistics for this new scenario can then be computed and are given in Table III.

Like in the previous section, we can then compare this to Table I in order to write the parameters s, c, and ϵ as functions of r and θ:

$$s = \tfrac{1}{2}(1+\cos\theta); \tag{A9}$$

$$c = (1-r)\sin^2\theta + \frac{r}{2}; \tag{A10}$$

$$\epsilon = \frac{r}{2}. \tag{A11}$$

Note that c and ϵ are the same as in the depolarizing case, but that s is now independent of r.

TABLE II. Statistics of the prepare-and-measure scenario with projectors depolarized by a parameter r.

Born rule	$\|\phi\rangle\langle\phi\|$	$\|\psi\rangle\langle\psi\|$	$\|\bar{\phi}\rangle\langle\bar{\phi}\|$	$\|\bar{\psi}\rangle\langle\bar{\psi}\|$
$\mathcal{D}_r^{\text{depol}}(\|\phi\rangle\langle\phi\|)$	$1-\frac{r}{2}$	$(1-r)\sin^2\theta+\frac{r}{2}$	$\frac{r}{2}$	$(1-r)\cos^2\theta+\frac{r}{2}$
$\mathcal{D}_r^{\text{depol}}(\|\psi\rangle\langle\psi\|)$	$(1-r)\sin^2\theta+\frac{r}{2}$	$1-\frac{r}{2}$	$(1-r)\cos^2\theta+\frac{r}{2}$	$\frac{r}{2}$
$\mathcal{D}_r^{\text{depol}}(E_{g_\phi})$	$\frac{1-r}{2}(1+\cos\theta)+\frac{r}{2}$	$\frac{1-r}{2}(1-\cos\theta)+\frac{r}{2}$	$\frac{1-r}{2}(1-\cos\theta)+\frac{r}{2}$	$\frac{1-r}{2}(1+\cos\theta)+\frac{r}{2}$

Finally, we can compute the maximal robustness to dephasing noise by demanding that inequality (A2) is saturated, that is, imposing the minimum dephasing noise $r_{\text{min}}^{\text{deph}}$ necessary for the existence of a noncontextual ontological model. This leads to the following equation for $r_{\text{min}}^{\text{deph}}$:

$$r_{\text{min}}^{\text{deph}} = 1 - \frac{1-\cos\theta}{\sin^2\theta}, \tag{A12}$$

which is exactly what we gave and discussed in the main text.

APPENDIX B: FAILURES OF NONCONTEXTUALITY CANNOT BE ACHIEVED WITHOUT SET COHERENCE

In this section, we give formal proof that one cannot prove the failure of the noncontextuality in scenarios where all the states (or all the effects) have no set coherence, that is, are simultaneously diagonalizable [44]. This means that computing robustness with respect to dephasing noise is a sensible measure of the failure of the existence of a noncontextual ontological model, as under sufficient dephasing noise *any* scenario will admit of a noncontextual ontological model. This result is well known in the community but we are not aware of an explicit proof so we include it here for convenience.

Proposition 1. Incoherent quantum states or measurements cannot prove the failure of noncontextuality.

Proof. We here give the proof for the case of incoherent states, with the case of incoherent measurements following straightforwardly. Consider a quantum system $\mathcal{H}$, a set of quantum states $\Omega := \{\rho_P\}_{P\in\mathcal{P}}$ for this system, and quantum effects $\mathcal{E} := \{E_{k|M}\}_{[k|M]\in K\times\mathcal{M}}$ acting on the system, such that $\sum_{k\in K} E_{k|M} = \mathbb{1}$, $\forall M \in \mathcal{M}$. Let $\{|i\rangle\}_{i\in I}$ be the basis in which all ρ_P are diagonalized, $I = \{0, 1, \dots, \dim\mathcal{H}\}$. We define a linear map $\mu : \Omega \to D[I] :: \rho_P \mapsto \mu_P$ where $D[I] \subset \mathbb{R}^I$ is the space of probability distributions over the index set I, where the μ_P are defined pointwise by

$$\mu_P(i) := \text{Tr}\{|i\rangle\langle i|\rho_P\} = \langle i|\rho_P|i\rangle, \quad \forall P \in \mathcal{P},\ \forall i \in I. \tag{B1}$$

These are indeed valid probability distributions as it is easy to show that $\sum_{i\in I} \mu_P(i) = 1$, $\forall P \in \mathcal{P}$. We also define a linear map $\xi : \mathcal{E} \to R[I] :: E_{k|M} \mapsto \xi_{k|M}$ where $R[I] \subset \mathbb{R}^I$ is the space of response functions over the index set I, where the $\xi_{k|M}$ are defined pointwise by

$$\begin{aligned} \xi_{k|M}(i) &:= \text{Tr}\{E_{k|M}|i\rangle\langle i|\} = \langle i|E_{k|M}|i\rangle, \\ &\forall [k|M] \in K\times\mathcal{M}, \quad \forall i \in I, \end{aligned} \tag{B2}$$

such that $\xi_{k|M}(i) \in [0, 1]\ \forall i \in I$ and $\sum_{k\in K} \xi_{k|M}(i) = 1\ \forall M \in \mathcal{M}$, $i \in I$, hence these constitute valid response functions.

Notice now that the quantum statistics in the experiment are reproduced by these probability distributions and response functions, since

$$\text{Tr}\{E_{k|M}\rho_P\} = \sum_{i\in I} \langle i|E_{k|M}\rho_P|i\rangle \tag{B3}$$

$$= \sum_{i,j\in I} \langle i|E_{k|M}|j\rangle\langle j|\rho_P|i\rangle \tag{B4}$$

$$= \sum_{i,j\in I} \langle i|E_{k|M}|j\rangle\langle i|\rho_P|i\rangle\delta_{ij} \tag{B5}$$

$$= \sum_{i\in I} \langle i|E_{k|M}|i\rangle\langle i|\rho_P|i\rangle \tag{B6}$$

$$= \sum_{i\in I} \xi_{k|M}(i)\mu_P(i), \tag{B7}$$

where for the third equality we used the fact that for all $P \in \mathcal{P}$, ρ_P is diagonal in the basis $\{|i\rangle\}_{i\in I}$. If we instead were working with incoherent measurements in this step, we would have instead used that $\langle i|E_{k|M}|j\rangle = \delta_{ij}\langle i|E_{k|M}|i\rangle$ in order to obtain the same result.

Finally, notice that whenever $\rho_P = \rho_{P'}$ (respectively, $E_{k|M} = E_{k'|M'}$), it will be the case that $\mu_P(i) = \mu_{P'}(i)$ [respectively, $\xi_{k|M}(i) = \xi_{k'|M'}(i)$], $\forall i \in I$, therefore constituting a noncontextual ontological model for the statistics of the scenario. ■

TABLE III. Statistics of the prepare-and-measure scenario, now with projectors dephased by a parameter r. Notice that the first two rows are the same as in the depolarized case (Table II). However, the entries in the last row do not depend on r since the dephasing noise does not change the effects aligned with the Z axis. Hence, the third row here is the same as the third row of Table II with r set to zero.

Born rule	$\|\phi\rangle\langle\phi\|$	$\|\psi\rangle\langle\psi\|$	$\|\bar{\phi}\rangle\langle\bar{\phi}\|$	$\|\bar{\psi}\rangle\langle\bar{\psi}\|$
$\mathcal{D}_r^{\text{deph}}(\|\phi\rangle\langle\phi\|)$	$1-\frac{r}{2}$	$(1-r)\sin^2\theta+\frac{r}{2}$	$\frac{r}{2}$	$(1-r)\cos^2\theta+\frac{r}{2}$
$\mathcal{D}_r^{\text{deph}}(\|\psi\rangle\langle\psi\|)$	$(1-r)\sin^2\theta+\frac{r}{2}$	$1-\frac{r}{2}$	$(1-r)\cos^2\theta+\frac{r}{2}$	$\frac{r}{2}$
$\mathcal{D}_r^{\text{deph}}(E_{g_\phi})$	$\frac{1}{2}(1+\cos\theta)$	$\frac{1}{2}(1-\cos\theta)$	$\frac{1}{2}(1-\cos\theta)$	$\frac{1}{2}(1+\cos\theta)$

APPENDIX C: ROBUSTNESS TO DEPHASING WITH THE LINEAR PROGRAM

1. Formal definitions

We begin this section by giving a formal definition of a GPT description of a given operational prepare-measure scenario [19,39].

Definition C.1. A *generalized probabilistic theory* associated with the operational scenario $(\mathcal{P}, \mathcal{M}, K, p)$ is a tuple $(V, \langle\cdot,\cdot\rangle, \Omega, \mathcal{E})$ such that

(1) $(V, \langle\cdot,\cdot\rangle)$ is a finite-dimensional, real vector space equipped with an inner product;

(2) $\Omega \subset V$ is a compact, convex set such that $V = \mathsf{LinSpan}[\Omega]$ and $0 \notin \mathsf{AffSpan}[\Omega]$, and where any element $s \in \Omega$, called a *state*, is associated with an inferential equivalence class of preparations, $\widetilde{P} \in \mathcal{P}/\simeq$;

(3) $\mathcal{E}$ is a subset of the dual Ω^*, such that both the origin 0 and the unit u (i.e., the unique vector satisfying $\langle u, s\rangle = 1$ for all $s \in \Omega$) in Ω^* are in $\mathcal{E}$, where any element $\varepsilon \in \mathcal{E}$ is called an *effect* and is associated with an inferential equivalence class of measurement outcomes, $\widetilde{[k|M]} \in \mathcal{M}/\simeq$; and

(4) for all $[k|M] \in \mathcal{M}$ and $P \in \mathcal{P}$, there is a respective $\varepsilon \in \mathcal{E}$ and $s \in \Omega$ such that

$$p(k|M, P) = \langle \varepsilon, s\rangle. \tag{C1}$$

A GPT is therefore a geometrical description of the operational scenario in which we have quotiented the sets of preparations and measurement outcomes by the inferential equivalence relation, since variations within the equivalence classes (i.e., the context of the procedure) are irrelevant for making predictions. In particular, this means that states and effects within the GPT satisfy the principle of *tomography*, i.e.,

$$\langle \varepsilon, s_1\rangle = \langle \varepsilon, s_2\rangle,\ \forall \varepsilon \in \mathcal{E} \iff s_1 = s_2;$$
$$\langle \varepsilon_1, s\rangle = \langle \varepsilon_2, s\rangle,\ \forall s \in \Omega \iff \varepsilon_1 = \varepsilon_2. \tag{C2}$$

The notion of nonclassicality employed in this work is the existence of a noncontextual ontological model of the operational scenario, which was shown in Ref. [19] to be equivalent to the simplex embeddability of the associated GPT. The latter is defined as follows.

Definition C.2. A GPT $(V, \langle\cdot,\cdot\rangle, \Omega, \mathcal{E})$ is *simplex embeddable* if and only if

(i) there exists $n \in \mathbb{N}$ defining

(1) the real vector space $\mathbb{R}^n$ with Euclidean inner product $\_ \cdot \_$, and

(2) the unit simplex $\Delta_n \in \mathbb{R}^n$ and its dual, the unit hypercube Δ_n^*,

(ii) there exists a pair of linear maps $\iota, \kappa : V \to \mathbb{R}^n$ such that

$$\iota(\Omega) \subseteq \Delta_n; \quad \kappa(\mathcal{E}) \subseteq \Delta_n^*, \tag{C3}$$

and (iii) the probabilistic predictions are preserved, i.e.,

$$\langle \varepsilon, s\rangle = \kappa(\varepsilon)\cdot \iota(s), \quad \forall s \in \Omega, \varepsilon \in \mathcal{E}. \tag{C4}$$

The simplex Δ_n can be thought of as the space of probability distributions over a finite set of ontic states, and the hypercube Δ_n^* as the response functions over the finite set. This can therefore be thought of as a geometric representation of the ontological theory in which we wish to represent the GPT. This ontological theory is formally equivalent to the GPT representation of classical probability theory [39].

Undeniably, it is not always the case that an experiment has access to *all* the states or effects in a GPT. In fact, in many cases the states and effects associated to an experiment will not even satisfy tomography. Moreover, if we consider nondeterministic sources as a way to prepare states in the experiment, then it can also be the case that subnormalized states can be prepared whilst their normalized counterparts cannot. An *accessible GPT fragment* of a GPT $(V, \langle\cdot,\cdot\rangle, \Omega, \mathcal{E})$ was defined in Ref. [27] to provide a description for such experiments. Formally it is a tuple $(I_{\Omega^F}, I_{\mathcal{E}^F}, \Omega^F, \mathcal{E}^F)$ such that $I_{\Omega^F}(\Omega^F) \subseteq \Omega$ and $I_{\mathcal{E}^F}^T(\mathcal{E}^F) \subseteq \mathcal{E}$, and where I_{Ω^F} and $I_{\mathcal{E}^F}$ are called *inclusion maps* [27]. Notice that there is no actual need to know the full GPT in order to define an accessible fragment—the elements in Ω^F and $\mathcal{E}^F$ can be written in terms of the subspaces they span (which might not be dual to each other) rather than with respect to the full vector space V. Due to this possible mismatch between the spanned spaces, the inclusion processes I_{Ω^F} and $I_{\mathcal{E}^F}$ are needed to provide the prediction rule, that is,

$$p(\epsilon, s) := \langle I_{\mathcal{E}^F}(\epsilon), I_{\Omega^F}(s)\rangle, \quad \forall \epsilon \in \mathcal{E}_F, s \in \Omega_F. \tag{C5}$$

The notion of a simplex embedding can be straightforwardly imported to the accessible GPT fragment, and the failure of simplex embeddability for a fragment immediately implies the nonexistence of an embedding for the full GPT [27,28]. Importantly for us, it has been shown in Ref. [28] that one can test for simplex embeddability using a linear program. Moreover, one can also use this linear program to compute how robust a given scenario is to depolarizing noise. In the following, we briefly introduce this linear program and show how it can be easily adapted to also compute robustness to dephasing noise.

2. Modification of the linear program

We begin this section by summarizing how the linear program from Ref. [28] works. Consider an accessible GPT fragment $(I_{\Omega^F}, I_{\mathcal{E}^F}, \Omega^F, \mathcal{E}^F)$. The linear program takes as inputs Ω^F and $\mathcal{E}^F$, and first characterizes the facet inequalities of the positive cones of states/effects. Suppose that there are $n_\Omega \in \mathbb{N}$ of these for states and $n_\mathcal{E} \in \mathbb{N}$ of these for effects. The linear program then turns these collections of facet inequalities into the matrices $H_\Omega : \mathsf{LinSpan}[\Omega^F] \to \mathbb{R}^{n_\Omega}$ and $H_\mathcal{E} : \mathsf{LinSpan}[\mathcal{E}^F] \to \mathbb{R}^{n_\mathcal{E}}$ such that

$$H_\Omega \cdot s \geqslant_e 0 \iff s = \sum_i q_i s_i, \quad s_i \in \Omega^F, q_i \in \mathbb{R}^+, \quad \forall i; \tag{C6}$$

$$H_\mathcal{E} \cdot \varepsilon \geqslant_e 0 \iff \varepsilon = \sum_i q_i \varepsilon_i, \quad \varepsilon_i \in \mathcal{E}^F, q_i \in \mathbb{R}^+, \quad \forall i, \tag{C7}$$

where $\geqslant_e$ is entry-wise non-negativity. The code also characterizes the inclusion map $I_{\Omega^F} : \mathsf{LinSpan}[\Omega^F] \to V$ (and $I_{\mathcal{E}^F} : \mathsf{LinSpan}[\mathcal{E}^F] \to V$) which maps each state (effect) from the accessible GPT fragment to the smallest Euclidean vector space V such that $\mathsf{LinSpan}[\Omega^F] \subseteq V$, $\mathsf{LinSpan}[\mathcal{E}^F] \subseteq V$ and with the dot product reproducing the probability rule. The code also takes as input a maximally mixed state $s_{\mathcal{D}}$, and characterizes the maximally depolarizing noise $\mathcal{D}$ from it. Finally, it solves the following linear program:

$$\begin{aligned} &\min \quad r \\ &\text{s.t.} \quad r I_{\mathcal{E}}^{T} \cdot \mathcal{D} \cdot I_{\Omega} + (1-r) I_{\mathcal{E}}^{T} \cdot I_{\Omega} = H_{\mathcal{E}}^{T} \cdot \sigma \cdot H_{\Omega}, \\ &\quad\quad \sigma \geqslant_e 0. \end{aligned} \tag{C8}$$

Now that we have summarized the main relevant aspect of the linear program of Ref. [28], we can introduce the particular accessible GPT fragment employed in our work. A pure qubit state $|\psi\rangle$ rotated from state $|0\rangle$ by an arbitrary angle θ about the Y axis can be represented in terms of an orthonormal operator basis as

$$|\psi\rangle\langle\psi| = \tfrac{1}{2}(\mathbb{1} + \sin\theta \hat{X} + \cos\theta \hat{Z}), \tag{C9}$$

where $\hat{X}$ (respectively, $\hat{Z}$) denotes the Pauli-X operator (respectively, Pauli-Z). We are assuming with no loss of generality that the plane in which our preparations and measurements live in the Bloch sphere is the ZX plane. Since Hermitian operators in this plane can be represented by a real-valued, three-dimensional vector, our states and effects will have the following form:

$$\boldsymbol{\psi} := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \sin\theta \\ \cos\theta \end{pmatrix}; \quad \bar{\boldsymbol{\psi}} := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -\sin\theta \\ -\cos\theta \end{pmatrix};$$

$$\boldsymbol{\phi} := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \sin\theta \\ -\cos\theta \end{pmatrix}; \tag{C10}$$

$$\bar{\boldsymbol{\phi}} := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -\sin\theta \\ \cos\theta \end{pmatrix}; \quad \boldsymbol{E}_{g_\psi} := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix};$$

$$\boldsymbol{E}_{g_\psi} := \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}. \tag{C11}$$

One can also define a null vector and a unit vector,

$$\mathbf{0} := \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}; \quad \boldsymbol{u} := \begin{pmatrix} \sqrt{2} \\ 0 \\ 0 \end{pmatrix}, \tag{C12}$$

and probabilities are given by taking the inner product between a preparation vector and an effect vector. This representation recovers all the expected statistics for this scenario. If we then define

$$\Omega^F := \mathrm{Conv}\{\psi, \bar{\psi}, \phi, \bar{\phi}\}, \tag{C13}$$

$$\mathcal{E}^F := \mathrm{Conv}\{\psi, \bar{\psi}, \phi, \bar{\phi}, E_{g_\psi}, E_{g_\phi}, 0, u\}, \tag{C14}$$

i.e., the convex hulls of the corresponding sets of vectors, then $\mathcal{F} := (\mathbb{1}, \mathbb{1}, \Omega^F, \mathcal{E}^F)$ is the accessible GPT fragment associated with the prepare-and-measure scenario studied in this work. In this scenario, we can take the inclusion maps to be identities because the states and effects of the scenario are mutually tomographic.

In quantum theory, a state ρ will dephase in the Bloch sphere when Z is the chosen basis, according to the dephasing channel $\mathcal{D}_Z$ defined by

$$\begin{aligned} \mathcal{D}_Z[\hat{\rho}] &:= \sum_{i\in\{0,1\}} \frac{1}{2}[\mathbb{1} + (-1)^i \hat{Z}]\hat{\rho}\frac{1}{2}[\mathbb{1} + (-1)^i \hat{Z}] \\ &= \frac{1}{2}(\mathbb{1} + \langle Z\rangle \hat{Z}). \end{aligned} \tag{C15}$$

In the representation of the scenario as an accessible GPT fragment $\mathcal{F}$, this dephasing channel is represented by the linear map $\mathcal{D}_Z : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\mathcal{D}_Z \circ \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \langle X\rangle \\ \langle Z\rangle \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ \langle Z\rangle \end{pmatrix}. \tag{C16}$$

More generally, for a general direction η in the ZX plane, we can define a dephasing map $\mathcal{D}_\eta$ in this representation. To start, define the projectors

$$\begin{aligned} |+_\eta\rangle\langle+_\eta| &= \tfrac{1}{2}(\mathbb{1} + \cos\eta\hat{X} + \sin\eta\hat{Z}), \\ |-_\eta\rangle\langle-_\eta| &= \tfrac{1}{2}(\mathbb{1} - \cos\eta\hat{X} - \sin\eta\hat{Z}), \end{aligned} \tag{C17}$$

from which it follows that

$$\begin{aligned} \mathcal{D}_\eta[\hat{\rho}] = \tfrac{1}{2}[&\mathbb{1} + (\langle X\rangle \cos^2\eta + \langle Z\rangle \cos\eta \sin\eta)\hat{X} \\ &+ (\langle X\rangle \cos\eta \sin\eta + \langle Z\rangle \sin^2\eta)\hat{Z}]. \end{aligned} \tag{C18}$$

For the accessible GPT fragment $\mathcal{F}$, this action of the dephasing map is hence represented by

$$\mathcal{D}_\eta \circ \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \langle X\rangle \\ \langle Z\rangle \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \langle X\rangle \cos^2\eta + \langle Z\rangle \sin\eta\, \cos\eta \\ \langle X\rangle \sin\eta\, \cos\eta + \langle Z\rangle \sin^2\eta \end{pmatrix}. \tag{C19}$$

This means that our GPT dephasing map in a generalized η basis (in the ZX plane) corresponds to the matrix

$$\mathcal{D}_\eta = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos^2\eta & \cos\eta\, \sin\eta \\ 0 & \cos\eta\, \sin\eta & \sin^2\eta \end{pmatrix}. \tag{C20}$$

The linear program from Ref. [28] takes as an input a set of states, a set of effects, and the unit vector from the accessible GPT fragment to be embedded. We modify the code to ask for an additional parameter η, and replace the occurrences of the depolarizing map with the matrix in Eq. (C20). This modification makes sense for the particular scenario that we are

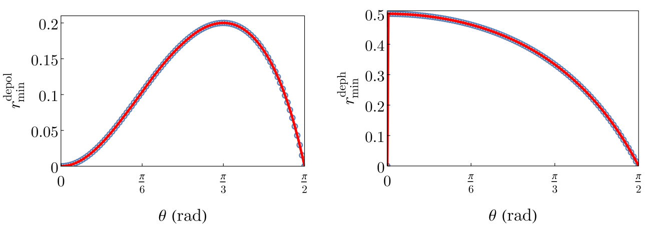


FIG. 5. To verify whether the original linear program (left) and its modification (right) survived the alterations, they were fed with the states and measurements of the original scenario (blue dots). Both obtained plots are compatible with the respective analytical plots (red curves).

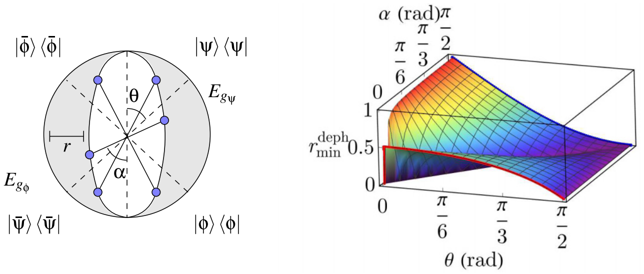


FIG. 6. Left: representation of the action of dephasing noise on the plane in which measurements live, now with measurement M_g rotated by an angle α. Right: contextual robustness to dephasing as a function of both the angle θ between the prepared states and the Z axis the angle α by which M_g is rotated. The red curve represents the plot from Fig. 3 of the main text and the blue curve represents the plot from Fig. 4.

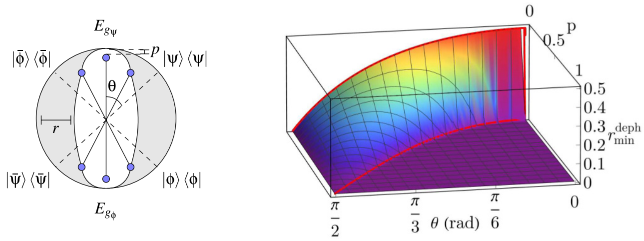


FIG. 7. Left: representation of the action of dephasing noise on the plane in which measurements live, now with discriminating measurement undergoing depolarizing noise by a factor p. Right: contextual robustness to dephasing as a function of both the angle between the prepared states and the Z axis and the noise added to the discriminating measurement. Red curves represent the plot from Fig. 3 of the main text and the equality from Eq. (D1).

interested in, but adapting the program for analyzing robustness to dephasing for general quantum scenarios and for other GPTs is beyond the scope of this work. Both the original linear program and the modified version also were altered to include the robustness as the first element of their string of outputs (at first, Ref. [28] would output only the list of epistemic states and response functions of the obtained noncontextual model), in order to make the plots easier. As a verification of whether the codes were functioning as expected, we demonstrated that both the original linear program and its modification managed to recover the analytical results from the main text, as shown in Fig. 5.

APPENDIX D: MESD WITH DEPHASING NOISE AND NOISY AND ROTATED DISCRIMINATING MEASUREMENT

Here we introduce the plots obtained via applying the modified linear program introduced in Appendix C to various scenarios (including different amounts and types of noise).

The first case we study is one where we rotate measurement M_g by a parameter α with respect to the Z axis, in order to investigate how the robustness to dephasing behaves in these related scenarios with a more coherent M_g. The plot is displayed in Fig. 6. As expected, the plot interpolates between the case in which M_g lies aligned to the Z axis (Fig. 3 of the main text) and the X axis (Fig. 4 of the main text). The only circumstances in which robustness is null are when the measurement M_g coincides with one of the other measurements, i.e., when $\alpha = \theta$. Furthermore, the plot shows clearly that $\alpha = \frac{\pi}{2}$ is the only circumstance in which the robustness saturates to 1.

The second case of study is the scenario where the measurement M_g has been affected by some amount p of depolarizing noise prior to the assessment of the robustness of the scenario to dephasing (i.e., before computing $r_{\min}$). The scheme for this scenario is given in Fig. 7, along with a plot for the robustness to dephasing as a function of both θ and p (the impurity of the discriminating measurements).

Notice that the section of the plot where $p = 0$ corresponds to the plot from Fig. 3 of the main text, that is, the dephased scenario with measurement M_g aligned with the Z axis. There is a clear relation between p and θ from which no contextuality can be proven, and numerically it coincides with the equation

$$p = 1 - \cos\theta. \tag{D1}$$

Equation (D1) hence provides the maximum noise one can add to the measurement M_g so there is still a proof of contextuality when the other measurements undergo dephasing noise. Geometrically, $p \geqslant 1 - \cos\theta$ represents a measurement M_g where the depolarizing noise p has made it such that its effects become merely convex combinations of the other effects in the scenario. Because one cannot prove contextuality with just the four preparations and their corresponding effects alone, scenarios with $p \geqslant 1 - \cos\theta$ admit of a noncontextual ontological model. If trace distance is the quantifier of coherence employed, as per Eq. (1) in the main text, then the inequality

$$C(|\psi\rangle\langle\psi|) > \sqrt{p(2-p)} \tag{D2}$$

tells us how much coherence the prepared states and measurements must start with so that the scenario can still prove contextuality despite the noise. Nevertheless, the scenario becomes more and more sensitive to dephasing noise as the impurity p increases, such that even for relatively small values of p the maximum robustness achieved decreases considerably. Notice still that for undisturbed measurement M_g ($p = 0$), inequality (D2) agrees with Result 1 of the main text: proofs of contextuality will be achieved as long $C(|\psi\rangle\langle\psi|) > 0$.

[1] D. Schmid, H. Du, J. H. Selby, and M. F. Pusey, Phys. Rev. Lett. **129**, 120403 (2022).
[2] D. Saha and A. Chaturvedi, Phys. Rev. A **100**, 022108 (2019).
[3] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Phys. Rev. Lett. **102**, 010401 (2009).
[4] A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai, Quantum Inf. Process **18**, 111 (2019).

[5] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, New J. Phys. **18**, 045003 (2016).
[6] S. A. Yadavalli and R. Kunjwal, Quantum **6**, 839 (2022).
[7] M. Lostaglio, Phys. Rev. Lett. **125**, 230603 (2020).
[8] M. Lostaglio and G. Senno, Quantum **4**, 258 (2020).
[9] D. Schmid and R. W. Spekkens, Phys. Rev. X **8**, 011015 (2018).
[10] J. Shin, D. Ha, and Y. Kwon, Entropy **23**, 1583 (2021).
[11] K. Flatt, H. Lee, C. R. I. Carceller, J. B. Brask, and J. Bae, PRX Quantum **3**, 030337 (2022).
[12] S. Mukherjee, S. Naonit, and A. K. Pan, Phys. Rev. A **106**, 012216 (2022).
[13] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).
[14] R. W. Spekkens, arXiv:1909.04628.
[15] D. Schmid, J. H. Selby, and R. W. Spekkens, arXiv:2009.03297.
[16] L. Catani, M. Leifer, D. Schmid, and R. W. Spekkens, arXiv:2111.13727.
[17] R. W. Spekkens, Phys. Rev. Lett. **101**, 020401 (2008).
[18] D. Schmid, J. H. Selby, M. F. Pusey, and R. W. Spekkens, arXiv:2005.07161.
[19] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, PRX Quantum **2**, 010331 (2021).
[20] R. D. Baldijão, R. Wagner, C. Duarte, B. Amaral, and M. T. Cunha, PRX Quantum **2**, 030351 (2021).
[21] D. Schmid, arXiv:2209.11783.
[22] Y. C. Liang, R. W. Spekkens, and H. M. Wiseman, Phys. Rep. **506**, 1 (2011).
[23] D. Schmid, R. W. Spekkens, and E. Wolfe, Phys. Rev. A **97**, 062103 (2018).
[24] V. J. Wright and M. Farkas, arXiv:2211.12550.
[25] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Phys. Rev. A **100**, 042116 (2019).
[26] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Phys. Rev. Lett. **130**, 230201 (2023).
[27] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Phys. Rev. A **107**, 062203 (2023).
[28] J. H. Selby, E. Wolfe, D. Schmid, and A. B. Sainz, arXiv:2204.11905.
[29] R. Kunjwal and R. W. Spekkens, Phys. Rev. Lett. **115**, 110403 (2015).
[30] I. Marvian, arXiv:2003.05984.
[31] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).
[32] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A **86**, 012103 (2012).
[33] R. W. Spekkens, in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer, Dordrecht, 2016), Vol. 181.
[34] W. H. Zurek, Rev. Mod. Phys. **75**, 715 (2003).
[35] M. Schlosshauer, arXiv:1404.2635.
[36] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).
[37] L. Hardy, arXiv:quant-ph/0101012.
[38] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[39] P. Janotta and H. Hinrichsen, J. Phys. A: Math. Theor. **47**, 323001 (2014).
[40] C. W. Helstrom, J. Stat. Phys. **1**, 231 (1969).
[41] A. Streltsov, G. Adesso, and M. B. Plenio, Rev. Mod. Phys. **89**, 041003 (2017).
[42] R. Wagner, R. S. Barbosa, and E. F. Galvão, arXiv:2209.02670.
[43] E. F. Galvão and D. J. Brod, Phys. Rev. A **101**, 062110 (2020).
[44] S. Designolle, R. Uola, K. Luoma, and N. Brunner, Phys. Rev. Lett. **126**, 220404 (2021).
[45] R. Wagner, A. Camillini, and E. F. Galvão, arXiv:2210.05624.

Robustness of contextuality under different types of noise as quantifiers for parity-oblivious multiplexing tasks

Amanda M. Fonseca,[1] Vinicius P. Rossi,[2,*] Roberto D. Baldijão,[2] John H. Selby,[2] and Ana Belén Sainz[2,3]

[1]*Institute of Physics, University of São Paulo, 05508-090 São Paulo, Brazil*
[2]*International Centre for Theory of Quantum Technologies, University of Gdańsk, 80-309 Gdańsk, Poland*
[3]*Basic Research Community for Physics e.V., Germany*

Generalised contextuality is a well-motivated notion of nonclassicality powering up a myriad of quantum tasks, among which is the celebrated case of a two-party information processing task where classical information is compressed in a quantum channel, the parity-oblivious multiplexing (POM) task. The success rate is the standard quantifier of resourcefulness for this task, while robustness-based quantifiers are as operationally motivated and have known general properties. In this work, we leverage analytical and numerical tools to estimate robustness of contextuality in POM scenarios under different types of noise. We conclude that for the 3-to-1 case robustness of contextuality to depolarisation, as well as a minimisation of robustness of contextuality to dephasing over all bases, are good quantifiers for the nonclassical advantage of this scenario. Moreover, we obtain a general relation between robustness of contextuality to depolarisation and the success rate in any n-to-1 POM scenario and show how it can be used to bound the number of bits encoded in this task.

Since the advent of quantum technologies, the superiority of quantum over classical methods has been anticipated across various areas, such as computation, information processing, and metrology. A celebrated result by Holevo [1] suggests that communication tasks may be an exception, stating that one can never retrieve more than $\log n$ bits of classical information out of n qubits, apparently forbidding quantum superiority over classical strategies for communicating. Quantum advantage, however, can arise in communication tasks that profit from the quantum systems' capacity to *carry* more information than classical ones [2, 3]. Therefore, such tasks prove interesting from both practical and foundational perspectives: they point to advantageous technological opportunities using quantum systems and also showcase how Holevo's theorem can be conciliated with such an advantage in communication. We here focus on one of such tasks, called *parity-oblivious multiplexing* (POM), in which Alice sends Bob an n-bit string by sharing a smaller number of qubits. It has been demonstrated that the maximum success probability using quantum strategies for POM tasks is strictly higher than the maximum success probability obtained using classical systems, constituting an instance of quantum advantage over classical strategies for communication [4].

It was previously shown that quantum advantage in POM tasks is powered by contextuality [5], one of the best-motivated notions of nonclassicality available [6]. Besides POM, contextuality powers up other aspects of communication [7–9], as well as computation [10], machine learning [11], information processing [12–15], metrology [16], state-dependent cloning [17], and state discrimination [18–21]. Contextuality first appeared in the literature in the Kochen-Specker theorem [22], and only recently it has been preliminary employed to make statements about generalised probabilistic theories (GPTs) [23–25] other than quantum theory. Generalisations of this notion have been developed in recent years, and Spekkens introduced an alternative approach that extends the notion of contextuality to encompass preparations, transformations, and unsharp measurements [5], contrasting with the Kochen-Specker theorem that primarily focuses on projective measurements. Since then, generalised contextuality has been proved to subsume or be related to many common notions of nonclassicality [4, 10, 18, 26–33], while challenging the nonclassical status of some phenomena [34–38]. Within the generalized contextuality framework – and therefore throughout this manuscript – a theory or scenario therein is deemed classical when it can be explained by a generalized noncontextual ontological model.

Proving whether a general operational scenario is classical is not straightforward, but Ref. [39] presents a linear program for testing this in any arbitrary prepare-and-measure scenario and a ready-to-use implementation available in Mathematica and Python[1]. Formally, to check for the existence of such a noncontextual model the code instead checks for the condition of *simplex- embeddability*, an equivalent notion of nonclassicality devised for GPTs [40]. Moreover, the code estimates how much depolarising noise must act on the experiment so that it admits of a noncontextual explanation, and it can be modified to include other noise models since it is open-source. This novel tool has already been employed to explore prepare-and-measure scenarios related to the quantum minimum-error state-discrimination protocol [41] and to provide a certification of contextuality in an experimental implementation of the quantum interrogation task [42].

Since we have good and operationally motivated measures for contextuality – robustness against different kinds of noise – and we know that contextuality powers up the advantage behind parity-oblivious multiplexing tasks, it is natural to ask how these contextuality measures relate to the natural quantifier in POM tasks, i.e., the success probability itself. In this work, we turn to this problem by exploring the generality allowed by the code in Ref. [39] to define different kinds of

* vinicius.prettirossi@phdstud.ug.edu.pl

[1] Available in Mathematica at https://github.com/eliewolfe/SimplexEmbedding and in Python at https://github.com/pjcavalcanti/SimplexEmbeddingGPT.

noise and compare its behaviour to the success probability in POM tasks. We find a general expression relating the robustness of contextuality to depolarisation and the success rate for n-to-1 POM tasks and show how it can be used to recover a well-known bound on the number of bits optimally encoded in a qubit in this task. Then, building up on Ref. [41], we generalise its modification to estimate robustness of contextuality to dephasing noise in a family of scenarios related to the 3-to-1 POM. Our numerical results show that robustness of contextuality to dephasing can also perform as a quantifier when minimised over specific axes. Moreover, since the code essentially tests for simplex-embeddability of noisy states and effects, the method we adopt here can also be used for testing a class of generalized probabilistic theories broader than just quantum theory.

I. PRELIMINARY CONCEPTS

A. Noncontextuality

The so-called operational approach is a way to construct physical theories based on laboratory experience [23, 24, 43]. This means that instead of the physical properties of a system and their time evolution, operational theories focus on how this physical system can be prepared, manipulated, and investigated in general experimental setups. Therefore, this approach is broad in the sense that it can be used to describe different kinds of theories, quantum and classical theories included. A typical *prepare-and-measure scenario* is described in this framework by the possible preparation procedures $P \in \mathcal{P}$, measurement procedures $M \in \mathcal{M}$, and possible outcomes $k \in K$ for each measurement that can be implemented in the experiment, as well as the statistics $p(k|M,P)$ obtained by implementing these procedures. Moreover, these scenarios are also characterised by equivalences between different procedures: two preparations, denoted as P and P', are deemed *operationally equivalent*, denoted as $P \simeq P'$, if they cannot be distinguished empirically, even in principle. This implies that any possible measurement outcome yields the same probability for both preparations. Similarly, two measurement outcomes, denoted as $k|M$ and $k'|M'$, are considered operationally equivalent, denoted as $k|M \simeq k'|M'$, if they both yield the same probability for any possible preparation. The tuple $(\mathcal{P}, \mathcal{M}, K, p, \simeq)$ defines an *operational theory*. As previously mentioned, quantum theory is one possible operational theory.

Generalized contextuality arises within this operational framework as the impossibility of an operational theory to conform with a noncontextual ontological model. An ontological model for an operational theory provides an underlying explanation for its statistics based on classical probability theory and Boolean logic, such that the behaviours captured by the operational theory can be reasoned about in classical terms. In particular, the ontological model maps the system to an ontic space Λ, preparation procedures are associated with probability distributions $\mu_P(\lambda)$, and measurement outcomes are associated with response functions $\xi_{k|M}(\lambda)$ over the ontic space, with $\lambda \in \Lambda$. This model explains the operational theory when, for all $P \in \mathcal{P}$ and $k|M \in K \times \mathcal{M}$, it satisfies

$$p(k|M,P) = \sum_{\lambda \in \Lambda} \xi_{k|M}(\lambda)\mu_P(\lambda). \tag{1}$$

Admitting of some ontological model is not yet a classical feature but a general one. Indeed, classicality has additional demands: any equivalences between operational procedures must hold at the ontological level[2], which underpins the notion of noncontextuality [5, 6]. An ontological model satisfies the assumption of noncontextuality if, for any $P, P' \in \mathcal{P}$, $k|M, k'|M' \in K \times \mathcal{M}$,

$$P \simeq P' \implies \mu_P(\lambda) = \mu_{P'}(\lambda), \quad \forall \lambda \in \Lambda, \tag{2}$$

$$k|M \simeq k'|M' \implies \xi_{k|M}(\lambda) = \xi_{k'|M'}(\lambda), \quad \forall \lambda \in \Lambda. \tag{3}$$

The condition of admitting of some noncontextual ontological model has proven to be a precise definition of classicality since the noncontextual ontological model itself provides a classical explanation to the empirical predictions. As was shown before, quantum theory does not admit such explanations [5], featuring nonclassicality.

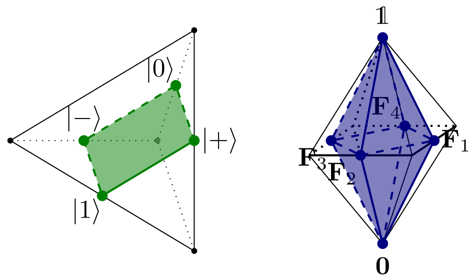
B. Operational measures of contextuality

Often, assessing noncontextuality of an operational scenario requires the characterisation of a large ontic space[3] and the possible epistemic states and response functions over it that can explain the statistical data and conform to the assumption of nonlocality for all equivalence relations. To avoid these issues when assessing the nonclassicality of the examples investigated in this work, we will employ the linear program introduced in Ref. [39], which relies on results from the framework of generalised probabilistic theories (GPT) [23–25]. We will not enter into the details of how the operational scenario translates to the GPT description, but a summary of the linear program and the reasoning behind it is provided in Appendix A. The relevant aspect for the present work is that this linear program decides the existence of a simplex embedding for the GPT fragment (see Fig. 1), which in turn is equivalent to the existence of a noncontextual ontological model for the associated operational scenario [40]. If the program cannot find a simplex embedding for the input scenario, i.e., the scenario exhibits contextuality, it calculates how much depolarising noise r is necessary so that the simplex embedding becomes possible. The implementation of the program takes in sets of states and effects of a GPT and outputs this minimal value of r along with the respective noncontextual ontological model explaining the scenario partially depolarised by a factor r.

---

[2] Otherwise, further arguments must be provided on why the procedures are not operationally distinguishable, despite being ontologically different.

[3] In quantum scenarios, for instance, the ontic space might be as large as $|\Lambda| = (\dim \mathcal{H})^2$ [39].



(a) Embedding of states (b) Embedding of effects

Figure 1. **Example of simplex embedding.** In the figure we see a qualitative example of a simplex-embedding: consider a scenario with the four pure qubit states in the equator of the Bloch ball depicted in (a) (and their convex combinations) and consider the effects to be the projection onto the same states – represented as $\{F_i\}_{i=1}^4$ in (b) – together with the 0 and $\mathbb{1}$ elements of POVM (and convex combinations thereof). This fragment can be represented as the green square in (a) and blue diamond-shaped object in (b). One can see that these objects can be embedded (colloquially, 'fit inside') into a classical GPT: the green set of states is embedded into a four-dimensional simplex (the tetrahedron in (a)) and the effects are embedded into the dual of the tetrahedron (the cube in (b)). If this is not possible to do – for simplices in any dimension – then there is no simplex embedding. For any GPT that admits no simplex-embedding, one can see that by shrinking either the effects or the states via some noise process, one will always make this procedure possible.

The quantity r hence may serve as an operational measure of nonclassicality, which we denote as the *robustness of contextuality to depolarisation*. It is known from simplex embedding reasoning that there is always a finite amount of partial depolarising noise under which an operational scenario becomes noncontextual, and explicit examples of this fact have been reported [18, 44, 45]. Alternatively, robustness can be defined in terms of different kinds of noise besides depolarising, so in this work, we also consider the case of dephasing. This is because it is a well-established fact that a completely dephased prepare-and-measure scenario will always admit of a noncontextual model [41], so the action of dephasing noise over a scenario will also eventually allow for a simplex embedding[4]. Therefore, we have further modified the code to estimate the robustness of contextuality to dephasing noise. Our modification works for any strongly self-dual GPT with finite dimension, and the reasoning behind it is given in detail in Appendix B.

C. Parity-oblivious multiplexing

In a n-to-1 parity-oblivious multiplexing (POM) task, a bitstring x of n classical bits is encoded in the preparation P_x of a quantum state. Measurements are constructed such that a binary measurement M_y returns outcome $k=0$ when $x_y=0$, where x_y is the y-th entry of the string x, and $k=1$ otherwise. Moreover, the encoding must be done so that no information about the parity of any subset of the string with more than one bit can be recovered from a single measurement.

Mathematically, one can define the parity of subsets of x as follows: consider the parity bitstrings $\{t|t=(t_1,\ldots,t_n),\quad \sum_i t_i \geq 2\}$; then, $x\cdot t$ tell us the parity of some subset of x determined by t – and the condition $\sum t_i \geq 2$ ensures this subset has at least two bits. Therefore, the set $\{x\cdot t\}_t$ encodes the parity information of all subsets of x containing more than one bit. With this notation, we say that a scenario satisfies parity-obliviousness if, given the parity strings $\{t\}$, it satisfies

$$\sum_{x|x\cdot t=0} p(k=x_y|M_y,P_x) = \sum_{x|x\cdot t=1} p(k=x_y|M_y,P_x), \qquad (4)$$

for all t, x and y.

In the quantum realisation of a 3-to-1 parity-oblivious task that is going to be explored in this work, a 3-bitstring is encoded in different preparations of a qubit. Parity-obliviousness imposes that mixing preparations with equal parity must yield the same statistical mixture as mixing preparations with the opposite parity. This adds symmetries to the shape formed by the convex hull formed by the preparations. We exploit these symmetries by considering implementations whose set of states is given by a rectangular cuboid, with each vertex parameterized by the angle θ that all preparations equally form with respect to the Z axis of the Bloch sphere. Measurements are assigned as the tomographic complete set $\hat{X},\hat{Y},\hat{Z}$ of Pauli measurements. This realisation obeys parity obliviousness constraints given in Eq. (4), and a schematic representation of this scenario can be found in Figure 2. Notice that even though Bob cannot retrieve more than 1 bit of information *in a single-shot*, he can *choose* which information to retrieve (though not deterministically). This can only be done because, even though one can at most retrieve 1 bit of information from a qubit, the qubit can carry more information than that. In this sense, Holevo's theorem is obeyed, but communication advantage is also achieved.

In general, the resourcefulness of a POM task is usually quantified by the probability of getting the outcome for measurement y that correctly matches the y-th bit in the string x, for any outcome and string, i.e.,

$$s = \frac{1}{2^n n}\sum_{y=0}^{n}\sum_{x\in\{0,1\}^{\times n}} p(b=x_y|M_y,P_x). \qquad (5)$$

It has been demonstrated that indeed the success rate is a good quantifier for a resource theory of contextuality [46]. Moreover, it is known that quantum realisations of this protocol can exceed the noncontextual bound of $s_{NC}=\frac{1}{2}\left(1+\frac{1}{n}\right)$ [12], attesting that contextuality is the source of quantum advantage for these tasks, since quantum realisations can achieve success rates of up to $s_Q=\frac{1}{2}\left(1+\frac{1}{\sqrt{n}}\right)$ [14], constituting instances of quantum advan-

[4] Although in this case, partial dephasing noise might not be enough to allow for the simplex embedding, as shown in Ref. [41].

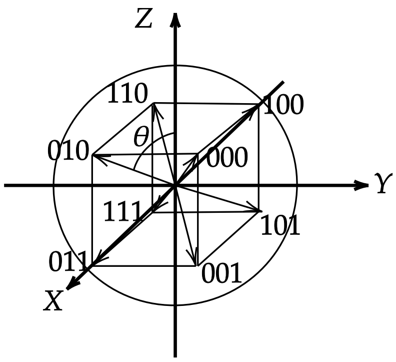


Figure 2. Representation of the preparation and measurement procedures of a 3-to-1 POM scenario among the family of scenarios investigated. Measurements lay on the X, Y, and Z axes, while preparations are represented by the vertices of the hexahedron.

tage for communication.

II. RESULTS

A. Robustness to depolarising noise in n-to-1 parity-oblivious multiplexing

Given that the advantage in POM tasks is underpinned by contextuality, and that depolarizing noise may transform a contextual scenario into a noncontextual one, it is natural to ask how much noise a POM scenario must receive to perform as poorly as classical realisations thereof. In particular, in the case of partial depolarising noise in a quantum realisation, every state under the action of this noise will be described by the mixture

$$\rho_x \mapsto (1-r)\rho_x + r\mu, \tag{6}$$

where μ is the maximally mixed state, and $r \in [0,1]$ is the noise quantifier. Notice that this represents a modification in all probabilities obtained in the scenario, since

$$\begin{aligned} p(b=x_y|P_x,M_y) &= \mathrm{Tr}(E_{b=x_y}\rho_x) \qquad (7)\\ &\mapsto (1-r)\,\mathrm{Tr}(E_{b=x_y}\rho_x) + r\,\mathrm{Tr}(E_{b=x_y}\mu)\\ &= (1-r)\,p(b=x_y|P_x,M_y) + r\,\mathrm{Tr}(E_{b=x_y}\mu), \end{aligned}$$

where $E_{b=x_y}$ are the POVM elements of measurement M_y. Once we calculate the success rate for the depolarised version of the experiment, we find

$$\begin{aligned} s_{depol} &= \frac{(1-r)}{2^n n}\sum_{y=1}^{n}\sum_{x\in[0,1]^{\times n}} p(b=x_y|P_x,M_y)\\ &\quad + \frac{r}{2^n n}\sum_{y=1}^{n}\sum_{x\in[0,1]^{\times n}} \mathrm{Tr}(E_{b=x_y}\mu) \qquad (8)\\ &= (1-r)s + \frac{r}{2n}\sum_{y=1}^{n}\mathrm{Tr}(\mu) \qquad (9)\\ &= (1-r)s + \frac{r}{2}, \qquad (10) \end{aligned}$$

where the second equality follows from the fact that for every choice of measurement y, there will be 2^{n-1} strings x with y-th input equal to 0, and 2^{n-1} strings with the same input equal to 1. Since all $E_{b=x_y}$ are elements of POVM, it must be that $E_{b=0_y} + E_{b=1_y} = \mathbb{1}$ for any y, simplifying the summation.

Let us define the *robustness of contextuality to depolarisation*, r_{min}^{depol}, as the minimum partial depolarising noise to be added to a scenario such that its depolarised version performs as poorly as a noncontextual realisation. In other words, r_{min}^{depol} is the minimum depolarizing noise such that the success rate of the depolarised scenario is the optimal noncontextual one, s_{NC}. Substituting this condition into Eq. (10), we get

$$r_{min}^{depol} = \frac{s - s_{NC}}{s - \frac{1}{2}}. \tag{11}$$

This proves the following result:

Proposition 1. In any n-to-1 parity-oblivious multiplexing scenario, success rate and robustness of contextuality to depolarisation are monotonically equivalent resource quantifiers via Eq. (11).

Proposition 1 then links a geometrical feature of states and effects used to implement a POM task and their success in such a task: the higher the depolarizing noise needed to make states P_x and measurements M_y become simplex-embeddable, the best they perform in the POM task, and vice-versa.

With this relation, one can derive bounds on the number of bits that can be optimally encoded in a particular quantum system by assuming bounds on the maximal robustness to depolarising achievable for this system. For instance, for qubits, it is common to assume that r_{min}^{depol} can never exceed the value $\frac{1}{2}$ for whatever signature of quantum advantage taken into consideration, in particular $r_{min}^{depol} < \frac{1}{2}$ for a finite amount of extremal states and effects in the fragment (which is always the case in n-to-1 POM tasks with finite n). We provide a compelling numerical induction for this bound in Appendix C 1 that relies on the simplex embeddability of an ever-increasing set of preparations and measurements. Then notice that the optimal quantum success rate is given by $s = \frac{1}{2}(1+\frac{1}{\sqrt{n}})$ [14], while $s_{NC} = \frac{1}{2}(1+\frac{1}{n})$ [4]. If $r_{min}^{depol} < \frac{1}{2}$, then

$$\begin{aligned} \frac{1}{2} &> \frac{s - s_{NC}}{s - \frac{1}{2}} \qquad (12)\\ &= \frac{\frac{1}{\sqrt{n}} - \frac{1}{n}}{\frac{1}{\sqrt{n}}} \qquad (13)\\ &= 1 - \frac{1}{\sqrt{n}}, \qquad (14) \end{aligned}$$

which means that $n < 4$. In other words:

Proposition 2. A two-dimensional quantum system used in a n-to-1 POM task can achieve a maximal advantage over classical systems only if $n < 4$.

This result has been derived before by geometrical [47] and locality [48] arguments, but Proposition 1 allows us to reframe it solely on simplicial embedding assumptions. This method can, in principle, be generalized: if one finds bounds on robustness for quantum systems of higher dimensions, one can immediately use Eq. (11) to derive an upper bound to the number of optimally encoded bits.

The linear program from Ref. [39] can also be used to get numerical results for different quantum implementations of an n-to-1 POM task. We focus on the case of a qubit and, given Prop. 2, the most interesting cases are $n = 2$ or $n = 3$. We will, however, focus on POM scenarios encoding 3 bits in a qubit, as introduced in Sec. I C. We deem these examples as more engaging than the simplest 2-to-1 POM since the latter is equivalent to the simplest prepare-and-measure scenario for which contextuality can be demonstrated, and much is already known about it [31, 49–52].

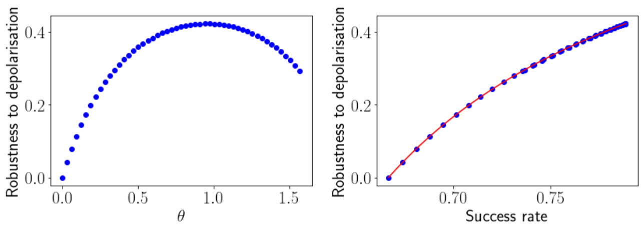


Figure 3. Robustness of contextuality to depolarising noise vs. parameter θ and vs. success rate for the 3-to-1 POM task. The red curve represents the plot of Eq. 11 for the 3-to-1 task.

In order to showcase this relation between robustness to depolarisation and success rate in a concrete example, we input the family of scenarios introduced in Sec. I C to the linear program, obtaining the plots in Figure 3. Notice that, as dictated by Prop. 1, robustness of contextuality to depolarising noise grows monotonically with the success rate. In particular, when $n = 3$, we have $s_{NC} = \frac{2}{3}$, and the numerical plot reproduces Eq. (11). We can also see that the peak on robustness of contextuality is achieved when $\theta = \frac{\pi}{3}$, i.e. when the states form a perfect cube inside the Bloch sphere. This scenario is compatible with the case in Ref. [12] for which the success rate is optimal. Intuitively, this can be related to the fact that this set of states has the greatest volume among the family of scenarios, and this situation would be expected to need the most noise for a simplex embedding to exist.

B. Robustness to dephasing noise in 3-to-1 parity oblivious multiplexing

Robustness-based quantifiers are advantageous in general since there is a fair knowledge about their properties, for instance, that in many resource theories, they are monotonic under linear combinations of free operations [53]. Moreover, robustness has a strong operational appeal since it relates to the ubiquitous characteristic in laboratories of uncontrolled degrees of freedom, and hence it would be interesting if statements about robustness of contextuality to other types of noise could be produced for the parity-oblivious tasks. Dephasing noise, for instance, is universally present in quantum computation and represents a considerable obstacle to the implementation of many protocols, especially concerning scalability. Moreover, it displays similarities with depolarising noise, in the sense that complete dephasing will deem any prepare-and-measure scenario noncontextual [41]. In this section, we thus analyze robustness to dephasing and how this contextuality measure relates to quantifying advantage in POM tasks. However, general results such as Prop. 1 are challenging when it comes to dephasing since it is a more structured type of noise, making arguments of this type difficult to build. We will therefore rely on the numerical investigation of the proposed 3-to-1 POM scenarios to investigate the usefulness of robustness to dephasing in quantifying the nonclassical advantage for these tasks by employing the modified code (see Appendix B).

For practical purposes, we chose the $\hat{Z}$ axis as the preserved axis, although the symmetry of this scenario should imply similar results for the $\hat{X}$ and $\hat{Y}$ axes. The results of our numerical explorations are presented in Fig. 4. As was previously shown in Ref. [41], the quantum advantage gets increasingly robust to dephasing as the preparations overlap with the dephasing basis. This however is not a good quantifier of this quantum advantage, since for the particular case in which robustness of contextuality is maximal, the success rate is very close to the classical bound, and robustness of contextuality decreases for scenarios in which the success rate improves. Since achieving a success rate $s > s_{NC}$ is the only contextuality inequality in this scenario, we can safely deem robustness to dephasing with respect to the $\hat{Z}$ axis as a bad quantifier[5]. It is interesting to see that, differently from the depolarising case, there are scenarios achieving a better than classical success rate that can endure a noise larger than 0.5.

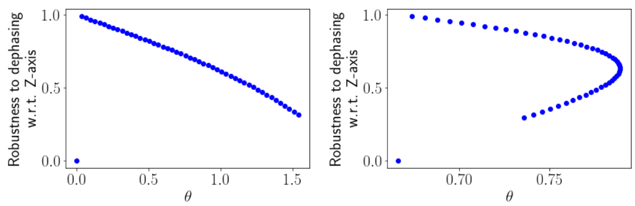


Figure 4. Robustness of contextuality to dephasing noise vs. parameter θ and vs. success rate for the 3-to-1 POM task.

Notice that the statement that this is not a good quantifier tells about how well the value of r captures the resourcefulness of the scenario for this particular task in contrast to

[5] The reason why this is the unique contextuality inequality is that $r > 0$ (be it robustness to depolarising or dephasing) happens if and only if there is no noncontextual ontological model for the scenario, and from Prop. 1 this is equivalent to having $s > s_{NC}$.

a classical realisation. However, robustness, in this case, is still a good *certifier* of nonclassicality, since as long as it is nonzero, it attests to the impossibility of a noncontextual ontological model for the scenario. This is possibly related to the axial symmetry of dephasing noise, when compared to the radial symmetry of the scenarios under investigation. Dephasing noise with respect to a particular axis is therefore not in a good fit with the symmetry suggested by the task, which motivates our following result.

We further show that robustness to dephasing can still perform as a good quantifier by exploiting the symmetry of this scenario and minimizing the robustness of contextuality to dephasing over all relevant dephasing bases. In particular, we here consider all available measurement basis $\hat{X}, \hat{Y}, \hat{Z}$. To gain intuition on why, notice that these axes will maximise the action of the noise over the volume of the set of states, requiring a smaller amount of noise to produce a greater change. Allowing for axes that align with a particular preparation, for instance, would possibly yield a larger value of robustness since the polyhedron from Fig. 2 would keep one of its diagonals unchanged, and the value of robustness with respect to this axis would be discarded by the minimisation. Indeed, we show the numerical investigation leading to this choice in Appendix C 2.

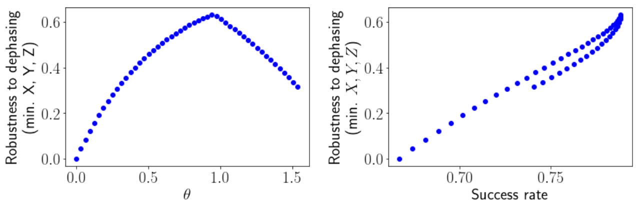


Figure 5. Minimal robustness of contextuality to dephasing noise across all of Bob's measurements vs. parameter θ and vs. success rate for the 3-to-1 POM task.

The plots are given in Figure 5. Maximal robustness of contextuality is now not achieved for small values of θ since, for the same scenarios, there is another dephasing axis with respect to which robustness of contextuality is almost null. In particular, dephasing with respect to either $\hat{X}$ or $\hat{Y}$ axes is the minimal one up until the parameter θ reaches $\theta = \frac{\pi}{3}$, precisely the scenario to which success rate is maximal. From that point on, the minimal robustness to dephasing is the one with respect to the $\hat{Z}$ axis.

The reason why this is still a good quantifier is that the extremal cases in which contextuality (and therefore success rate) is minimal and maximal are captured as extremal cases by the minimised robustness to dephasing, i.e., it assigns $r = 0$ only for $s = s_N$, and its highest value coincides only with $s = s_Q$. Evidently, this minimised robustness distinguishes scenarios that the success rate does not and vice-versa, so we cannot deem them as monotonically proportional. However, specifying a value of r in most cases singles out the value of s, and in the worst scenario it is upper and lower bounded by the success rate. These bounds do increase monotonically when they exist, which displays some sort of coherence between these two quantifiers [54].

It is interesting to see how this quantity displays a coherent relation with the success rate similar to the depolarising case, even though robustness of contextuality is not minimised over all possible axes in the Bloch sphere. It would be interesting to investigate whether this quantifier, when taken with respect to all possible axes, has any relation to quantities that are Bargmann invariants such as basis-independent coherence [55, 56].

III. CLOSING REMARKS

Our work builds up on the results from Refs. [39, 41] to further explore the role of different types of noise over the possibility of a noncontextual ontological model in prepare-and-measure scenarios. We derive an analytical relation between robustness of contextuality to depolarisation and the success rate that usually quantifies the quantum advantage in POM tasks (Prop. 1). We then use it to recover a result that has been previously derived from both geometrical [47] and locality arguments [48] that a POM with optimal success rate is impossible when encoding more than 3 bits in a qubit. In particular, we give a re-framing of these proofs solely on simplex-embedding arguments, and we believe that more general statements can be made by observing the relation between robustness of contextuality and success rate, increasing the insight provided by these proofs.

Furthermore, we present a generalisation of the linear program from Ref. [39] estimating robustness of contextuality to dephasing noise for any GPT fragment of a strongly self-dual GPT, with quantum theory as the example of interest, and leverage it to explore the particular case of the 3-to-1 parity-oblivious multiplexing task, to which contextuality is known to be a resource. We conclude that beyond robustness of contextuality to depolarisation, other noise models (such as a minimisation of robustness of contextuality to dephasing over a set of bases) can also display a consistent relation with respect to the standard resource quantifier. In contrast, robustness to dephasing with respect to the $\hat{Z}$ axis does not display such a relation. It however remains a good certifier of nonclassical advantage, since for any nonzero value of robustness, advantage over classical implementations is ensured. These results showcase the versatility of the linear program introduced in Ref. [39] since many experimental constraints can be taken into consideration when deciding the nonclassicality of a prepare-and-measure scenario.

It would be interesting to investigate how the quantifiers explored in our framework relate to quantifiers of Bargmann invariants [55, 56] and quantifiers obtained by SDP hierarchies [57, 58].

Another direction of future research includes a better understanding of how noise models beyond depolarisation should be defined in more complex GPTs, for instance, in weakly self-dual GPTs. The operational approach to quantum resource theories has gained space in the literature [46, 59–61] and might provide useful insights on the relation between different signatures of nonclassicality.

ACKNOWLEDGEMENTS

We thank Bárbara Amaral for all the support and guidance on this manuscript, and Rafael Wagner, Giulio Camillo, and Anubhav Chaturvedi for the fruitful discussion. A.M.F. was supported by the São Paulo Research Foundation (FAPESP) via Project Numbers 2022/05455-5 (FAS-Masters) and 2022/13208-8 (RIA-Research Internship Abroad), and from the Instituto Serrapilheira, Grant 2021/4. V.P.R. was supported by the Foundation for Polish Science (IRAP project, ICTQT, contract no. MAB/2018/5, co-financed by EU within Smart Growth Operational Programme). J.H.S. was supported by the National Science Centre, Poland (Opus project, Categorical Foundations of the Non-Classicality of Nature, Project No. 2021/41/B/ST2/03149). R.D.B. and A.B.S. acknowledge support by the Digital Horizon Europe project FoQaCiA, Foundations of quantum computational advantage, GA No. 101070558, funded by the European Union, NSERC (Canada), and UKRI (UK). Some figures were prepared using TikZit.

---

[1] A. S. Holevo, Problems Inform. Transmission **9**, 177 (1973).
[2] N. Gisin and R. Thew, Nature Photonics **1**, 165–171 (2007).
[3] G. Brassard, Foundations of Physics **33**, 1593–1616 (2003).
[4] R. W. Spekkens, Phys. Rev. Lett. **101**, 020401 (2008).
[5] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).
[6] R. W. Spekkens, arXiv:1909.04628 [physics.hist-ph] (2019).
[7] D. Saha and A. Chaturvedi, Phys. Rev. A **100**, 022108 (2019).
[8] A. S. S., S. Mukherjee, and A. K. Pan, Phys. Rev. A **107**, 012411 (2023).
[9] C. Roch i Carceller, K. Flatt, H. Lee, J. Bae, and J. B. Brask, Phys. Rev. Lett. **129**, 050501 (2022).
[10] D. Schmid, H. Du, J. H. Selby, and M. F. Pusey, Phys. Rev. Lett. **129** (2022).
[11] J. Bowles, V. J. Wright, M. Farkas, N. Killoran, and M. Schuld, arXiv:2302.01365 [quant-ph] (2023).
[12] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Phys. Rev. Lett. **102**, 010401 (2009).
[13] A. Ambainis, M. Banik, A. Chaturvedi, D. Kravchenko, and A. Rai, arXiv:1607.05490 [quant-ph] (2016).
[14] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, New J. Phys. **18**, 045003 (2016).
[15] S. A. Yadavalli and R. Kunjwal, Quantum **6**, 839 (2022).
[16] M. Lostaglio, Phys. Rev. Lett. **125**, 230603 (2020).
[17] M. Lostaglio and G. Senno, Quantum **4**, 258 (2020).
[18] D. Schmid and R. W. Spekkens, Phys. Rev. X **8**, 011015 (2018).
[19] J. Shin, D. Ha, and Y. Kwon, Entropy **23**, 1583 (2021).
[20] K. Flatt, H. Lee, C. R. I. Carceller, J. B. Brask, and J. Bae, PRX Quantum **3**, 030337 (2022).
[21] S. Mukherjee, S. Naonit, and A. K. Pan, Phys. Rev. A **106**, 012216 (2022).
[22] S. Kochen and E. P. Specker, The Problem of Hidden Variables in Quantum Mechanics, in *The Logico-Algebraic Approach to Quantum Mechanics: Volume I: Historical Evolution*, edited by C. A. Hooker (Springer Netherlands, Dordrecht, 1975) pp. 293–328.
[23] L. Hardy, arXiv:quant-ph/0101012 (2001).
[24] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[25] P. Janotta and H. Hinrichsen, J. Phys. A: Math. Theor. **47**, 323001 (2014).
[26] D. Schmid, J. H. Selby, and R. W. Spekkens, arXiv:2009.03297 [quant-ph] (2020).
[27] R. D. Baldijão, R. Wagner, C. Duarte, B. Amaral, and M. T. Cunha, PRX Quantum **2**, 030351 (2021).
[28] Y. C. Liang, R. W. Spekkens, and H. M. Wiseman, Physics Reports **506**, 1 (2011).
[29] V. Wright and M. Farkas, arXiv:2211.12550 [quant-ph] (2022).
[30] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Phys. Rev. A **100**, 042116 (2019).
[31] L. Catani, M. Leifer, D. Schmid, and R. W. Spekkens, Quantum **7**, 1119 (2023).
[32] D. Schmid, R. D. Baldijão, Y. Yīng, R. Wagner, and J. H. Selby, arXiv:2405.04573 [quant-ph] (2024), 2405.04573 [quant-ph].
[33] J. Lin, K. Wang, L. Xiao, D. Qu, G. Zhu, Y. Zhang, and P. Xue, Phys. Rev. A **109**, 052208 (2024).
[34] L. Catani, M. Leifer, D. Schmid, and R. W. Spekkens, arXiv:2111.13727 [quant-ph] (2021).
[35] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, arXiv:2106.09045 [quant-ph] (2021).
[36] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).
[37] S. D. Barlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A **86**, 012103 (2012).
[38] R. W. Spekkens, in *Quantum Theory: Informational Foundations and Foils*, Vol. 181, edited by G. Chiribella and R. W. Spekkens (Springer, Dordrecht, 2016).
[39] J. H. Selby, E. Wolfe, D. Schmid, A. B. Sainz, and V. P. Rossi, Phys. Rev. Lett. **132**, 050202 (2024).
[40] D. Schmid, J. H. Selby, E. Wolfe, R. Kunjwal, and R. W. Spekkens, PRX Quantum **2**, 010331 (2021).
[41] V. P. Rossi, D. Schmid, J. H. Selby, and A. B. Sainz, Phys. Rev. A **108**, 032213 (2023).
[42] T. Giordani, R. Wagner, E. Esposito, A. Camillini, F. Hoch, G. Carvacho, C. Pentangelo, F. Ceccarelli, S. Piacentini, A. Crespi, N. Spagnolo, R. Osellame, E. F. Galvão, and F. Sciarrino, Science Advances **9**, eadj4249 (2023), https://www.science.org/doi/pdf/10.1126/sciadv.adj4249.
[43] G. M. D'Ariano, G. Chiribella, and P. Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, 2017).
[44] I. Marvian, arXiv:2003.05984 [quant-ph] (2020).
[45] R. Kunjwal and R. W. Spekkens, Phys. Rev. Lett. **115**, 110403 (2015).
[46] L. Catani, T. D. Galley, and T. Gonda, arXiv:2406.00717 [quant-ph] (2024).
[47] M. Hayashi, K. Iwama, H. Nishimura, R. Raymond, and S. Yamashita, New Journal of Physics **8**, 129 (2006).
[48] S. Ghorai and A. K. Pan, Phys. Rev. A **98**, 032110 (2018).
[49] M. F. Pusey, Phys. Rev. A **98**, 022112 (2018).
[50] M. Khoshbin, L. Catani, and M. Leifer, Phys. Rev. A **109**, 032212 (2024).
[51] L. Catani, M. Leifer, G. Scala, D. Schmid, and R. W. Spekkens, Phys. Rev. Lett. **129**, 240401 (2022).
[52] L. Catani, M. Leifer, G. Scala, D. Schmid, and R. W. Spekkens, Phys. Rev. A **108**, 022207 (2023).

[53] C. Duarte and B. Amaral, J. Math. Phys. , 062202 (2018).
[54] B. Regula, Journal of Physics A: Mathematical and Theoretical **51**, 045303 (2017).
[55] E. F. Galvão and D. J. Brod, Phys. Rev. A **101** (2020).
[56] R. Wagner, R. S. Barbosa, and E. F. Galvão, arXiv:2209.02670 [quant-ph] (2022).
[57] A. Chaturvedi, M. Farkas, and V. J. Wright, Quantum **5**, 484 (2021).
[58] G. Viola, T. Prasad, E. Panwar, and A. Chaturvedi, In preparation..
[59] J. H. Selby and C. M. Lee, Quantum **4**, 319 (2020).
[60] D. Schmid, D. Rosset, and F. Buscemi, Quantum **4**, 262 (2020).
[61] B. Zjawin, D. Schmid, M. J. Hoban, and A. B. Sainz, Quantum **7**, 1134 (2023).
[62] J. H. Selby, D. Schmid, E. Wolfe, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Phys. Rev. A **107**, 062203 (2023).
[63] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, New Journal of Physics **13**, 063024 (2011).
[64] M. P. Müller and C. Ududec, Phys. Rev. Lett. **108**, 130401 (2012).
[65] P. Janotta and R. Lal, Phys. Rev. A **87**, 052131 (2013).
[66] H. Barnum and J. Hilgert, arXiv.1904.03753 [math-ph] (2019).

Appendix A: Linear program for testing nonclassicality – Formal definitions

The linear program introduced in Ref. [39] is the main tool employed in this manuscript and is constructed entirely in the GPT framework. Here we give some formal definitions for this program and our further modification of what some of the notions mentioned in the main text comprise.

The notion of operational equivalence enables us to define a GPT for the associated operational theory. For this, we concentrate on the quotiented sets $\mathcal{P}/\simeq$ and $K\times\mathcal{M}/\simeq$, i.e., the sets of equivalence classes $[P]$ and $[k|M]$ such that

$$[P] := \{P' \in \mathcal{P} | P' \simeq P\}; \tag{A1}$$

$$[k|M] := \{k'|M' \in K\times\mathcal{M} | k'|M' \simeq k|M\}. \tag{A2}$$

In this framework, every equivalence class $[P] \in \mathcal{P}/\simeq$ in the quotiented operational theory is associated to a state $s_P \in \Omega$, and every equivalence class $[k|M] \in K\times\mathcal{M}/\simeq$ is associated to an effect $e_{k|M} \in \mathcal{E}$. Each state or effect can be described as a vector in a real vector space V equipped with an inner product $\langle\cdot,\cdot\rangle$. Additionally, these sets must satisfy some specific properties, the most important of which being that the sets Ω and $\mathcal{E}$ are tomographic. A GPT can be defined as a tuple $(V,\langle\cdot,\cdot\rangle,\Omega,\mathcal{E})$ satisfying these properties, and a formal definition for the purposes of this paper is given below:

Definition A.1 (GPT)**.** A *generalised probabilistic theory* associated to an operational scenario is a tuple $(\Omega,\mathcal{E},V,\langle\cdot,\cdot\rangle)$ where Ω and $\mathcal{E}$ are convex sets of vectors in a real vector space V equipped with an inner product $\langle\cdot,\cdot\rangle$. Moreover,

- Ω does not contain the origin $0 \in V$, and each element $s_P \in \Omega$ is associated with an equivalence class $[P] \in \mathcal{P}/\simeq$;
- $\mathcal{E}$ contains the origin and the privileged *unit* effect u, and each element $e_{k|M} \in \mathcal{E}$ is associated with an equivalence class $[k|M] \in K\times\mathcal{M}/\simeq$;
- Probabilities are given via the inner product, $p(k|M,P) = \langle s_P, e_{k|M}\rangle$;
- For all $s \in \Omega$, $\frac{1}{\langle s,u\rangle}s \in \Omega$, where u is the unit effect;
- Ω and $\mathcal{E}$ are *tomographic*, i.e.,

$$s_1 = s_2 \iff \langle s_1, e\rangle = \langle s_2, e\rangle, \quad \forall e \in \mathcal{E}; \tag{A3}$$

$$e_1 = e_2 \iff \langle s, e_1\rangle = \langle s, e_2\rangle, \quad \forall s \in \Omega. \tag{A4}$$

As mentioned in the main text, however, the linear program in Ref. [39] is written to also take into consideration scenarios in which some of the above properties are not satisfied. In principle, one can simply conceive a relaxation of a GPT in which one merely has access to subsets of the sets Ω and $\mathcal{E}$. This already implies dropping some features, such as the inclusion of normalised counterparts or tomographic completeness. Such a relaxation receives the name of *GPT fragment*.

Definition A.2 (GPT fragments)**.** A *GPT fragment* associated to an operational scenario is a tuple $(\Omega^F,\mathcal{E}^F,V,\langle\cdot,\cdot\rangle)$ such that $\Omega^F \subseteq \Omega$ and $\mathcal{E}^F \subseteq \mathcal{E}$, with $(\Omega,\mathcal{E},V,\langle\cdot,\cdot\rangle)$ being a GPT.

Alternatively, these states and effects can be naturally described with respect to the subspaces of V they span (which will most probably not match[6]), respectively $\mathsf{Span}(\Omega)$ and $\mathsf{Span}(\mathcal{E})$, and from here on we will always assume this is the case unless stated otherwise. One can easily describe these vectors with respect to the full vector space V by employing inclusion maps $I_\Omega : \mathsf{Span}(\Omega) \to V$ and $I_\mathcal{E} : \mathsf{Span}(\mathcal{E}) \to V$, in which case the tuple $(\Omega,\mathcal{E},I_\Omega,I_\mathcal{E})$ is referred to as the *accessible GPT fragment* [62].

Definition A.3 (Accessible GPT fragments)**.** An *accessible GPT fragment* associated with an operational scenario is a tuple $(\Omega^A,\mathcal{E}^A,I_\Omega,I_\mathcal{E})$, such that[7]

- $I_\Omega(\Omega^A) \in V$, $I_\mathcal{E}(\mathcal{E}^A) \in V$;

---

[6] Consider, for instance, the POM scenario introduced in Sec. I C for $\theta = \frac{\pi}{2}$. In this case, all states lie in the equator of the Bloch sphere, spanning $\mathbb{R}^2$, while the effects remain lying on the three Pauli axes and therefore spanning $\mathbb{R}^3$. This is only one of many examples in which the sets of states and effects in a fragment span different subspaces.

[7] In fact, accessible GPT fragments must satisfy more properties than these, for instance, the sets Ω^A and $\mathcal{E}^A$ must have more structure than merely convexity. We leave these nuances out of this manuscript since they are not relevant to any of the conclusions drawn.

- $p(k|M,P) = \langle I_\Omega(s), I_{\mathcal{E}}(e)\rangle$, for all $[P] \in P/\simeq$ and $[k|M] \in K \times \mathcal{M}/\simeq$;
- $(I_\Omega(\Omega^A), I_{\mathcal{E}}, (\mathcal{E}^A), V, \langle\cdot,\cdot\rangle)$ is a GPT fragment.

Beyond characterising these sets and the subspaces they span, we can further characterise their positive cones, i.e., the sets $\mathsf{Cone}(\Omega)$ of states that are linear combinations with non-negative coefficients of the states in Ω, and similarly for $\mathsf{Cone}(\mathcal{E})$. In th case that $(\Omega,\mathcal{E})$ are polytopes, we can characterise their cones by the facet inequalities $\{h_i^\Omega\}_{i=1}^n$ for states and $\{h_j^{\mathcal{E}}\}_{j=1}^m$ for effects, where $h_i^\Omega : \mathsf{span}(\Omega) \to \mathbb{R}$ and $h_j^{\mathcal{E}} : \mathsf{Span}(\mathcal{E}) \to \mathbb{R}$. These are linear maps such that

$$h_i^\Omega(v) \geq 0 \iff v \in \mathsf{Cone}(\Omega),\, i = 1,\dots,n, \tag{A5}$$

$$h_j^{\mathcal{E}}(w) \geq 0 \iff w \in \mathsf{Cone}(\mathcal{E}),\, j = 1,\dots,m. \tag{A6}$$

Equivalently, we can concatenate the sets $\{h_i^\Omega\}_{i=1}^n$ and $\{h_j^{\mathcal{E}}\}_{j=1}^m$ into matrices H_Ω, $H_{\mathcal{E}}$ such that

$$H_\Omega(v) := (h_1^\Omega(v),\dots,h_n^\Omega(v))^T, \forall v \in \mathsf{Span}(\Omega), \tag{A7}$$

$$H_{\mathcal{E}}(w) := (h_1^{\mathcal{E}}(w),\dots,h_n^{\mathcal{E}}(w))^T, \forall w \in \mathsf{Span}(\mathcal{E}). \tag{A8}$$

Notice that, by construction, $H_\Omega(v)$ is entry-wise non-negative iff $v \in \mathsf{Cone}(\Omega)$, and similarly for $H_{\mathcal{E}}$ iff $w \in \mathsf{Cone}(\mathcal{E})$. Using this characterisation of the accessible GPT fragment, we can introduce the linear program developed in Ref. [39].

Program 1. Let r be the minimum depolarising noise that must be added for the statistics obtained by composing any state-effect pair from Ω and $\mathcal{E}$ to be classically explainable. Given $(H_\Omega, H_{\mathcal{E}}, I_\Omega, I_{\mathcal{E}})$ characterising the cones of the accessible GPT fragment, r can be computed by the linear program:

minimize r such that

$$\exists\ \sigma \geq_e 0, \text{ an } m\times n \text{ matrix such that}$$

$$rI_{\mathcal{E}}^T \cdot D_{depol} \cdot I_\Omega + (1-r)I_{\mathcal{E}}^T \cdot I_\Omega = H_{\mathcal{E}}^T \cdot \sigma \cdot H_\Omega \tag{A9}$$

where D_{depol} is the completely depolarising channel for the system, and $\geq_e$ denotes entry-wise non-negativity.

Notice therefore that the linear program assesses the simplex-embeddability of accessible GPT fragments that are not necessarily within quantum theory. The notion of depolarising noise for such cases is in fact much broader since the code does not demand any particular property from the state labeled as maximally mixed. Abstractly, the code simply "shrinks" the set of states towards a point specified by the input maximally mixed state until a simplex embedding becomes possible, which in case of a qubit happens to be the center of the Bloch sphere.

This linear program decides nonclassicality because the existence of a matrix σ satisfying equation A9 is equivalent to the existence of a simplex embedding for the GPT fragment depolarised by a factor r. This, in turn, has been demonstrated to imply the existence of a noncontextual ontological model for the associated operational scenario [40]. The quantity r therefore may serve as an operational measure of nonclassicality, which we denote as the *robustness of contextuality*.

Appendix B: Modification of the code

The implementation for the linear program in Ref. [39] requests as inputs the sets of states and effects $(\Omega,\mathcal{E})$ that constitute the fragment, a privileged effect u that plays the role of the *discard* effect and a privileged state μ, called *maximally mixed* state. The code then constructs the accessible GPT fragment associated with the input fragment and the fully depolarising map D_{depol}, consisting of discarding any state it receives and preparing the maximally mixed state instead. The code in Mathematica has the additional feature of checking whether the input states and effects are represented by density operators and POVM element matrices, and if they do, already assumes the discard effect to be $\mathbb{1}_{\mathcal{H}}$ and the maximally mixed state to be $\mathbb{1}_{\mathcal{H}}/\dim(\mathcal{H})$, where $\mathcal{H}$ is the Hilbert space with respect to which the density operators and POVM elements are represented.[8] Both implementations then proceed to characterise the accessible GPT fragment and its cone facets and solve Program 1. Finally, they construct an ontological model for the depolarised scenario by factorising the matrix σ found by the program, and output an array $(r,\vec{\mu},\vec{\xi})$, where r is the robustness of contextuality to depolarisation, $\vec{\mu}$ is a list of epistemic states for the noncontextual ontological model, and $\vec{\xi}$ a list of the respective response functions.

Ref. [41] modified the original Mathematica code so that instead of requiring a maximally mixed state as input, it would take a real parameter $\eta \in [0,\pi)$ representing an angle between an axis in the ZX plane of the Bloch sphere and the Z axis. The depolarising map is then replaced by an explicit matrix representation of the fully dephasing map with respect to the axis singled out by η, and this modification is employed to investigate the interplay between contextuality and coherence in a family of prepare-and-measure scenarios related to the minimum-error state-discrimination task. Although successful for this particular task, this modification is rather limited, since it will not give an account of any GPT fragment whose associated accessible GPT fragment cannot be described in the hemisphere of the Bloch sphere. So even though many important tasks to which contextuality is a resource fit in the scope of this modification, we are interested in a more general approach that can cover the whole scope of quantum prepare-and-measure scenarios.

[8] In this case, the user does not need to provide the discard and the maximally mixed state.

In the modification introduced by this manuscript, the code asks for the GPT fragment $(\Omega^{\mathcal{F}}, \mathcal{E}^{\mathcal{F}})$, the discard effect, and a finite collection of effects $\mathcal{M} = \{e_i\}_{i=1}^{m}$, and then constructs the fully dephasing map as the process that realises the effects $e_i \in \mathcal{E}$ and prepare the respective states $\bar{e}_i \in \Omega$ for which $\langle \bar{e}_i, e_i \rangle = 1$ for all $i = 1, \ldots, m$, and where $(V, \langle \cdot, \cdot \rangle, \Omega, \mathcal{E})$ is the GPT from which $(\Omega^{\mathcal{F}}, \mathcal{E}^{\mathcal{F}})$ is derived. Notice therefore that this modification will not give an account of GPTs in which the effects characterising the dephasing axis do not have normalised states as counterparts. In practice, what the code does is to construct the matrix

$$D_{deph} = (\bar{e}_1 \ldots \bar{e}_m) \cdot (e_1 \ldots e_m)^T, \tag{B1}$$

such that $\bar{e}_i = e_i^T$, and which will replace the depolarising map D_{depol} in Equation A9.

Notice that such implementation is not valid for all possible GPTs, since to conform to Definition A.1 does not imply that all effects $\bar{e}_i$ have corresponding states $\bar{e}_i$ as introduced in the previous paragraph. In demanding this structure for the noise model we restrict ourselves to a subset of the *strongly self-dual GPTs* [63–65]. Given a GPT $(V, \langle \cdot, \cdot \rangle, \Omega, \mathcal{E})$, and their positive cones $\mathsf{Cone}(\Omega) \subset V$ and $\mathsf{Cone}(\mathcal{E}) \subset V$. Then:

Definition B.1 (Strongly self-dual GPT)**.** A GPT $(V, \langle \cdot, \cdot \rangle, \Omega, \mathcal{E})$ is *strongly self-dual* if there is a symmetric, positive-semidefinite isomorphism $T : \mathsf{Cone}(\mathcal{E}) \to \mathsf{Cone}(\Omega)$ such that, for any $e_1, e_2 \in \mathsf{Cone}(\mathcal{E})$,

- $\langle T(e_1), e_2 \rangle = \langle T(e_2), e_1 \rangle$;
- $\langle T(e_1), e_1 \rangle \geq 0$.

That is, strongly self-dual GPTs are those in which one can find a symmetric and positive-semidefinite T under which the cones of effects and states are isomorphic $\mathsf{Cone}(\mathcal{E}) \simeq \mathsf{Cone}(\Omega)$. Note that GPTs that are not strongly self-dual can be made turned into self-dual GPTs if one restricts the set of effects or states in a suitable way [65]. In particular, the self-dual theories considered in this work are the ones whose isomorphism T is a transposition of the effect vector. This might miss, for instance, GPTs associated to some Euclidian Jordan algebras [66].

Notice also that in principle there is no constraint over the set $\mathcal{M}$ of dephasing effects, for instance, it is not required to form a normalised measurement or to have orthogonal terms. As an additional constraint, the code also checks whether the effects in $\mathcal{M}$ are orthogonal to each other and sum up to the unit effect. This constraint makes sure that, for the quantum case, only the common notion of dephasing is going to be investigated, and these constraints by no means undermine the generality of our results since there is not a clear definition of how a dephasing noise model should look like in GPTs other than quantum theory, so we merely focus our attention to noise models that for sure match the standard notion of quantum dephasing noise. Investigating the noise models ruled out by these constraints, i.e., the ones in which the effects provided by $\mathcal{M}$ are incomplete or coarse-grained, representing projections onto hypervolumes other than an axis, or investigating whether GPTs that are not strongly self-dual accommodate some notion of dephasing noise, are interesting directions for future research. For instance, the Linear Program 1 and modifications thereof rely only on properties of the positive cones of states and effects to decide the nonclassicality of GPT fragments. It would be interesting to see if weakly self-dual GPTs[9] can accommodate some notion of dephasing that still leads to trustworthy statements of simplex-embeddability.

Another aspect of this choice of noise model that deserves attention is that, in principle, the set $\mathcal{M}$ that characterises the dephasing axis does not need to be a subset of $\mathcal{E}$, neither the state counterparts of these effects need to belong in Ω. This might lead to situations in which assessments of nonclassicality in these scenarios become inconclusive since the dephased GPT fragment might not admit of a simplex embedding for any value of r. We argue however that allowing for this sort of noise is a reasonable choice that captures the nature of noise in many experiments. Indeed, noise is usually understood as a process replacing the original information encoded in the system with undesired one, which might include information making the experiment more “nonclassical” than it is. Furthermore, this is also the approach taken in the formulation of the original Linear Program, in which the maximally mixed state provided is not required to be in Ω [39]. This can also lead to situations in which the whole set of states is “shrunk” towards a point that lies outside it, and yet it can capture instances of noise other than depolarisation, such as decay to a (pure) ground state.

Finally, as an example of our reasoning, consider a scenario that takes states and effects from a qubit and in which dephasing happens with respect to the Z axis. The code will receive as input the set $\mathcal{M} = \{\frac{\mathbb{1}+\hat{Z}}{\sqrt{2}}, \frac{\mathbb{1}-\hat{Z}}{\sqrt{2}}\}$, and the dephasing map takes the form

$$D_{deph} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{B2}$$

Naturally, this map will act on any state or effect such that

$$D_{deph} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} \langle \mathbb{1} \rangle \\ \langle \hat{X} \rangle \\ \langle \hat{Y} \rangle \\ \langle \hat{Z} \rangle \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \langle \mathbb{1} \rangle \\ 0 \\ 0 \\ \langle \hat{Z} \rangle \end{pmatrix}, \tag{B3}$$

which is a point in the Z axis of the Bloch sphere.

[9] Weakly self-dual GPTs are the ones whose positive cones of states and effects are isomorphic, with no further restrictions on the isomorphism linking them.

Appendix C: Numerical investigations

1. Numerical motivation for the bound $r < \frac{1}{2}$ for n-to-1 POM with a qubit

To build up evidence that justifies the bound $r \leq \frac{1}{2}$ employed in the main text, we once again rely on a numerical investigation using the linear program from Ref. [39]. For this purpose, we consider a prepare-and-measure scenario with an ever-increasing number of equally distributed preparations and measurement outcomes.

For a scenario with $2n$ preparations and measurement outcomes over a qubit Hilbert space, consider the sets

$$\Omega_n := \left\{ |k\rangle = \cos\left(\frac{k\pi}{2n}\right)|0\rangle + \sin\left(\frac{k\pi}{2n}\right)|1\rangle \right\} \tag{C1}$$

$$\mathcal{E}_n := \left\{ |k\rangle = \cos\left(\frac{k\pi}{2n} + \frac{\pi}{8}\right)|0\rangle + \sin\left(\frac{k\pi}{2n} + \frac{\pi}{8}\right)|1\rangle \right\} \tag{C2}$$

with $k = 0, \ldots, 2n$. For instance, for $n = 2$, this example comprises the preparations and measurement outcomes used in the optimal 2-to-1 POM task. Notice that, for $n \to \infty$, this comprises all preparations and measurements in the real hemisphere of the Bloch sphere. We input these sets to the code from Ref. [39] to observe the relation between robustness to depolarisation and the number $2n$ of preparations and outcomes. This plot is given in Figure 6.

It is immediate to see that robustness quickly approaches the value $\frac{1}{2}$, from bellow. This alone is enough to build motivation for expecting $r < \frac{1}{2}$ in a POM task encoding a finite amount of bits into a finite amount of preparations since these values will only be achievable for $n \gg 3$ preparations and measurements.

As argued in the main text, one can easily generalise this method of inspection to quantum systems of greater dimension, recovering bounds of robustness of contextuality and consequently, of optimal quantum performance for parity-oblivious multiplexing tasks via Eq. 11.

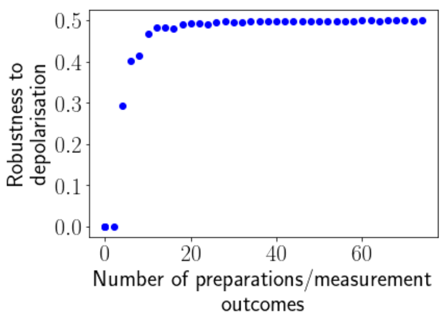


Figure 6. Robustness to depolarisation with the number of preparations and measurement outcomes in the real hemisphere of the Bloch sphere.

2. Numerical motivation for neglecting dephasing axes beyond X, Y and Z

In Sec. II B, we chose minimising robustness to dephasing over the axes X, Y, and Z. We here show by a numerical investigation that indeed the axes X, Y, and Z are the only ones that matter.

By using a pseudo-random number generator function, we can generate random axes of the form

$$\left\{ \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ \sin\theta_r\cos\phi_r \\ \sin\theta_r\sin\phi_r \\ \cos\theta_r \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -\sin\theta_r\cos\phi_r \\ -\sin\theta_r\sin\phi_r \\ -\cos\theta_r \end{pmatrix} \right\}, \tag{C3}$$

where θ_r and ϕ_r are independently generated random angles.

We then compute the relation between robustness to dephasing and success rate with respect to the Z axis; minimised over X, Y and Z axis; minimised over X, Y, Z and n random axes, with n being 2, 10, 20 and 100, respectively. The plot is displayed in Figure 7.

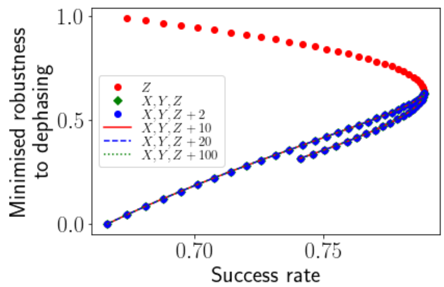


Figure 7. Robustness to dephasing vs. success rate for the 3-to-1 POM task, minimised over axes X, Y, Z and additional pseudo-random axes.

Notice that once we minimise robustness to dephasing over axes X, Y, and Z, the addition of the extra random axes becomes irrelevant. Indeed, all points for the plots with extra axes coincide with the plots without it. Despite this not being a minimisation over all possible axes, it provides enough evidence for our conjecture to hold.