Testing classical and quantum theory with single photons

by

Michael Mazurek

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics (Quantum Information)

Waterloo, Ontario, Canada, 2018

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Wolfgang Tittel Professor, University of Calgary
Supervisor	Kevin Resch Professor, University of Waterloo
Internal Members	Robert Spekkens Faculty, Perimeter Institute for Theoretical Physics Adjunct Faculty, University of Waterloo
	Kyung Soo Choi Assistant Professor, University of Waterloo
Internal-external Member	Joseph Emerson Associate Professor, University of Waterloo

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

This thesis is based on the following articles:

Chapters 3 and 5, and Appendix B contain material from [1]:

> 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," *arXiv:1710.05948* (2017).
>
> Robert Spekkens conceived of the experiment.
>
> Michael Mazurek and Kevin Resch designed the experiment.
>
> Michael Mazurek performed the experiment and collected the data.
>
> All authors discussed how to interpret the data and contributed to the final draft of the manuscript.
>
> Michael Mazurek and Matthew Pusey implemented the numerical data analysis.
>
> Michael Mazurek and Robert Spekkens wrote the initial draft of the manuscript.

Chapter 4 and Appendix A contain material from [2]:

> M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch and R. W. Spekkens, "An experimental test of noncontextuality without unphysical idealizations" *Nature Communications*, **7**:11780 (2016).
>
> Ravi Kunjwal and Rob Spekkens proved the noncontextuality inequality.
>
> Michael Mazurek and Kevin Resch designed the experiment and implemented the numerical analysis of the data.
>
> Michael Mazurek performed the experiment and showed that the bound of the noncontextuality inequality was tight.

Matthew Pusey developed the method to overcome the problem of inexact operational equivalence.

All authors contributed to writing the manuscript.

Contributions not included in this thesis:

J. M. Donohue, M. D. Mazurek and K. J. Resch "Theory of high-efficiency sum-frequency generation for single-photon waveform conversion," *Phys. Rev. A*, **91**, 033809 (2015). [3]

X. Ma, T. Jackson, H. Zhou, J. Chen, D. Lu, M. D. Mazurek, K. A. G. Fisher, X. Peng, D. Kribs, K. J. Resch, Z. Ji, B. Zeng, and R. Laflamme, "Pure-state tomography with the expectation value of Pauli operators," *Phys. Rev. A*, **93**, 032140 (2016). [4]

Abstract

To date, quantum theory is the most successful physical theory that has been discovered. However, there is still the possibility that an inconsistency between experimental observations and the predictions of quantum theory may one day be found, thus prompting the replacement of quantum theory with a superior, post-quantum theory. To narrow down the scope of possibilities for the true theory that describes nature, one can perform experiments that falsify other physical theories, by demonstrating an incompatibility between experimental observations and the predictions of these theories. This thesis details two such experiments. First we provide relevant background information, beginning with a review of experimental quantum optics. Next, we review noncontextual ontological theories and discuss requirements for experimental tests of such theories. Finally, we discuss the framework of generalised probabilistic theories before introducing the two experiments.

The first experiment is a test of noncontextual ontological (or hidden-variable) models of nature. An ontological model of a physical theory is one in which systems have preexisting properties, and a noncontextual ontological model is one in which systems that are indistinguishable experimentally are represented identically in the model. Physical theories that cannot be represented by a noncontextual ontological model are said to be nonclassical; quantum theory is an example of such a physical theory that is nonclassical in this sense. Prior to this thesis, experimental tests of the assumption of noncontextuality had assumed that the experiments were free of both systematic and statistical errors, which is not justifiable for any experiment. We introduce new analytical techniques that allow us to avoid making these assumptions, and perform an experiment with single photons that, with high confidence, rules out the possibility of describing nature with a noncontextual ontological model.

The second experiment is a demonstration of self-consistent state and measurement tomography in the framework of generalised probabilistic theories (GPTs). The GPT framework is a very general, operationally-motivated framework for describing a physical theory in terms of the observable events predicted by the theory. We develop a technique for inferring the GPT description of a set of states and measurements directly from experimental data. By analysing our data in this general framework, we are able to test various candidate physical theories of nature. We perform an experiment with single photons, and quantify the size of possible variations between quantum theory and the true physical theory that describes nature.

Acknowledgements

I owe thanks to many individuals who have helped make my time in Waterloo as great as it was.

I'll start by thanking my supervisor, Kevin Resch. Thank you for your open door policy, your extremely constructive comments on so many practice talks, your insight when it comes to solving problems in the lab, and your excitement to talk about science. Thank you to Rob Spekkens for your incredible skill in scientific reasoning, and for agreeing to join my examing committee at the last minute. Thank you to my committee members Joseph Emerson, Kyung Soo Choi, and Donna Strickland for your guidance and advice. Thank you to external examiner Wolfgang Tittel.

John Donohue and Kent Fisher deserve special thanks for participating in vine-growing competitions, egg-fountain bets, candy time, and fruitful discussions. Thank you John for your trivia, karaoke, and deep-fryer skills, and Kent for your wordsmithery and love of office sports. Thank you Jean-Philippe MacLean for helping me out with Blender and for trying to teach me about wine. Thank you Jeff Salvail for being my coolest friend. Thank you Maggie Agnew or teaching me LabVIEW. Thank you to Deny Hamel for building (and to Lydia Vermeyden for improving upon) a great photon source. Thank you to Jonathan Lavoie and Robert Prevedel for teaching me the basics of how to work in the lab. Thank you to other current and past members of the QOQI group: Matt Brown, Patrick Daley, Morgan Mastrovich, Andrew Cameron, and Krister Shalm. Thank you also to Aimee Gunther, Chris Pugh, and Sascha Agne for sharing optics.

Thank you to Matt Pusey and Ravi Kunjwal for teaching me much of what I know about quantum foundations, and for showing me where the ping pong table is at PI.

Thank you to the members of the UW OSA student chapter for deciding to build a museum exhibit, and thank you to Donna Strickland for encouraging us to do it. Thank you Angela Olano for the insane amount of work you did to help us make it happen, and thanks to both you and Meg Hawk for your love of brunch.

Thank you Gary Graham and Olivia Di Matteo for looking after Frank when we were out of town, and thank you Gary for looking after Gina.

Thank you above all else to my family: Corey Rae, Cheryl, Dave, Liz, Cormac, Marni, Ken, and Molly.

To Corey Rae.

Table of Contents

List of Tables

List of Figures

Chapter 1

Experimental quantum optics

Experimental quantum optics is the study of light at the single or few-photon level. The experiments presented in Chapters 4 and 5 are performed using single photons that are produced via a nonlinear process called spontaneous parametric downconversion. In this chapter we present the background necessary to understand the experiments presented in later chapters, within the framework of quantum theory.

We begin in Section 1.1 by introducing the definitions of quantum states, transformations, and measurements. The quantization of the electric field is reviewed in Section 1.2, and creation of photon pairs via spontaneous parametric downconversion is reviewed in Section 1.3. Section 1.4 describes how the polarization degree-of-freedom of a single photon can be used to encode quantum information, and also how this information can be prepared, manipulated and measured in experiments. We end in Section 1.5 with descriptions of quantum state, process, and measurement tomography; these are procedures for inferring the quantum operators describing state preparation, transformation, and measurement devices from experimental data.

1.1 Quantum states, transformations, and measurements

The quantum description of an experiment can be broken down into three main components: system preparation, transformation, and measurement. In this section we review each of these components.

1.1.1 Quantum states

A pure quantum state of a physical system is an element of a complex vector space on which we can define an inner product [5]. We call such a vector space a *Hilbert space*. The set of all possible states of the system is called the *state space*.

We denote a pure quantum state with $|\psi\rangle$, and for the purposes of this thesis $|\psi\rangle$ represents a d-dimensional column vector with complex coefficients, where d is the dimension of the Hilbert space. The conjugate transpose of $|\psi\rangle$ is denoted with $\langle\psi|$. A pure state $|\psi\rangle$ is a unit vector in Hilbert space, which means the *inner product* of $|\psi\rangle$ with itself, $\langle\psi \mid \psi\rangle$, is equal to 1. The *dimension* of the quantum system, d, is the maximum number of states that can be perfectly distinguished from each other with a single-shot measurement. Some examples of sets of distinguishable states for different physical systems are the discrete atomic energy levels an electron can occupy, the $+1/2$ and $-1/2$ spin states of a neutron, or the Hermite-Gauss spatial modes of a photon. We can denote a set of d distinguishable states as $\{|0\rangle, \ldots, |d-1\rangle\}$. Any two distinguishable states are orthogonal, meaning that $\langle i \mid j\rangle = \delta_{i,j}$, where $\delta_{i,j}$ is the Kronecker delta. Thus, a set of d distinguishable states forms an orthonormal basis for the d-dimensional state space.

In order to represent a general quantum state (i.e., a state that is not necessarily pure), one must use a *density matrix*. The density matrix, ρ, representing a pure state $|\psi\rangle$ is the *outer product* of $|\psi\rangle$ with itself, that is, $\rho = |\psi\rangle\langle\psi|$. The state space is convex, and pure states lie on the boundary of the space. A convex combination of two or more pure states is a mixed state, and is represented by a convex combination of density operators. For example, consider two preparation devices: one can prepare the pure state $|\psi\rangle$, and the other can prepare $|\phi\rangle$. Now consider a third preparation device that flips a weighted coin that lands heads with probability w and tails with probability $1-w$. If the coin lands heads, the state $|\psi\rangle$ is prepared, and if the coin lands tails then $|\phi\rangle$ is prepared. The state ρ prepared by this third preparation device is represented by the convex combination of the states produced by the first two devices, that is $\rho = w|\psi\rangle\langle\psi| + (1-w)|\phi\rangle\langle\phi|$.

Every density operator has to be positive semidefinite[1] (i.e. its eigenvalues are all greater than or equal to zero), denoted by $\rho \geq 0$. If we are free to choose the definition of the orthonormal basis $\{|0\rangle, \ldots, |d-1\rangle\}$, then the most general density operator in d-dimensional Hilbert space, is written as:

$$\rho = \sum_{i=0}^{d-1} w_i |i\rangle\langle i|, \tag{1.1}$$

[1] In this thesis, when referring to operators, we will use the terms "positive" and "positive semidefinite" interchangeably.

which must satisfy $\sum_i w_i = 1$ and $w_i \geq 0$ since each w_i represents a probability. From this definition one can see that ρ is Hermitian ($\rho = \rho^\dagger$, where † denotes a conjugate transpose) and has trace equal to one ($\mathrm{Tr}\,\rho = 1$).

The decomposition of a density matrix (in the form of Eq. (1.1)) is not necessarily unique. For example, consider the density operator ρ_{mixed}, which is an equal mixture of two pure states, $\rho_{\text{mixed}} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$. We can define the rotated basis $\{|+\rangle, |-\rangle\}$, where $|\pm\rangle = \frac{|0\rangle \pm |1\rangle}{\sqrt{2}}$, and represent our state as $\rho_{\text{mixed}} = \frac{1}{2}|+\rangle\langle +| + \frac{1}{2}|-\rangle\langle -|$.

1.1.2 Transformations of quantum states

A general state transformation Λ takes a state ρ to another valid state $\rho' = \Lambda(\rho)$. For example, a pure state $|\psi\rangle$ can be transformed to another pure state $|\psi'\rangle$ via a unitary operator U, $|\psi'\rangle = U|\psi\rangle$ [5]. If $\rho = |\psi\rangle\langle\psi|$ and $\rho' = |\psi'\rangle\langle\psi'|$, then evolution from $|\psi\rangle$ to $|\psi'\rangle$ in the density operator representation is $\rho' = U\rho U^\dagger$.

It is also possible to take convex mixtures of unitary evolutions. For example, one could consider, with probability w, transforming a state ρ with unitary U, and with probability $1 - w$, transforming the state with unitary V. In this case the resulting state ρ' is, in general, mixed, and equal to:

$$\rho' = wU\rho U^\dagger + (1-w)V\rho V^\dagger. \tag{1.2}$$

The most general form of a state transformation Λ is [5]:

$$\Lambda(\rho) = \sum_i K_i \rho K_i^\dagger \tag{1.3}$$

This is called the *Kraus representation* of Λ, and the set of operators $\{K_i\}$, are called the *Kraus operators*. The Kraus operators are $d \times d$, and they do not have to be unitary or Hermitian. However, the full set must satisfy:

$$\sum_i K_i^\dagger K_i = \mathbb{I}, \tag{1.4}$$

which, along with the fact that $\mathrm{Tr}\,\rho = 1$ and $\rho \geq 0$, ensures that $\Lambda(\rho)$ is positive and trace one [5] ($\mathbb{I}$ denotes the identity operator).

1.1.3 Quantum measurements

A K-outcome quantum measurement, M, on a d-dimensional quantum system is represented with a set of $d \times d$ positive operators $\{E_0, \ldots, E_{K-1}\}$ with the outcomes $\{0, \ldots, K-1\}$ assigned to each operator, respectively [5]. We call each E_k a *measurement effect*, and we call the measurement M a *positive operator-valued measure* (POVM) (and sometimes we will refer to the full set of effects $\{E_0, \ldots, E_{K-1}\}$ as a POVM as well). If a system in state ρ is measured, the probability of getting outcome k is $p(k|M, \rho) = \mathrm{Tr}(\rho E_k)$.

Probabilities are positive, real numbers, (i.e. $\mathrm{Tr}(\rho E_k) \geq 0$ for any valid density operator ρ), which implies that measurement effects must be positive operators: $E_k \geq 0$. Also, since the probability of the measurement returning *any* outcome is one, we must have $\sum_{i=0}^{K-1} \mathrm{Tr}(\rho E_k) = 1$ which in turn implies $\sum_{i=0}^{K-1} E_k = \mathbb{I}$.

A special type of quantum measurement is a *projective valued measure* (PVM). A PVM in d-dimensional Hilbert space is represented with a set of $k \leq d$ effects, $\{\Pi_0, \ldots, \Pi_{k-1}\}$, satisfying $\Pi_i^2 = \Pi_i$, for $i \in \{0, \ldots, k-1\}$. A *rank-1* PVM in d-dimensional Hilbert space is represented with a set of d rank-1 effects, $\{\Pi_0, \ldots, \Pi_{d-1}\}$, with $P_k = |k\rangle \langle k|$ for some choice of orthogonal basis $\{|0\rangle, \ldots, |d-1\rangle\}$. A rank-1 PVM can reliably distinguish between states in this basis in a single-shot measurement, because $\mathrm{Tr}(\Pi_k |l\rangle \langle l|) = \delta_{k,l}$.

For the sake of notational simplicity above, we labelled the outcomes of the measurement M with $\{0, \ldots, K-1\}$, but of course we have the freedom to label the outcomes in any way we want. Thus, we can label the outcomes of M as $\{m(k)\}$, where outcome $m(k)$ corresponds to effect E_k.

For an input state ρ the expectation value of a measurement, M, is calculated with the formula $\sum_{k=0}^{K-1} m(k) \mathrm{Tr}(\rho E_k) = \mathrm{Tr}\left(\rho \sum_{k=0}^{K-1} m(k) E_k\right)$. We define the *observable* for measurement M with $O_M = \sum_{k=0}^{K-1} m(k) E_k$, which allows us to calculate the expected value of measurement M as $\langle O_M \rangle = \mathrm{Tr}(\rho O_M)$.

1.1.4 The qubit

We will illustrate the above ideas using the $d = 2$ dimensional Hilbert space as an example. A quantum system described by such a space is called a *qubit*. A pure qubit state is written in the form $\alpha |0\rangle + \beta |1\rangle$, where α and β are both complex numbers that satisfy $|\alpha|^2 + |\beta|^2 = 1$.

A general 2×2 Hermitian operator can be written as a linear combination of the 2×2

identity operator, $\mathbb{I}$, and the 2×2 *Pauli operators* σ_x, σ_y, and σ_z [6], where

$$\begin{aligned} \sigma_x &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \\ \sigma_y &= \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ \sigma_z &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned} \tag{1.5}$$

The Pauli operators are traceless, and they have eigenvalues $+1$ and -1. We define the *computational basis*, $\{|0\rangle, |1\rangle\}$ as the eigenstates of σ_z. $|0\rangle$ is the eigenstate with eigenvalue $+1$, and $|1\rangle$ is the eigenstate with eigenvalue -1. The $+1$ eigenstate of σ_x is $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and the -1 eigenstate is $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$. The $+1$ eigenstate of σ_y is $|+i\rangle = \frac{1}{\sqrt{2}}(|0\rangle + i\,|1\rangle)$ and the -1 eigenstate is $|-i\rangle = \frac{1}{\sqrt{2}}(|0\rangle - i\,|1\rangle)$.

Defining $\boldsymbol{\sigma} = \begin{pmatrix} \mathbb{I} & \sigma_x & \sigma_y & \sigma_z \end{pmatrix}$ we can write a general trace-one Hermitian operator as $\rho = \frac{1}{2}\boldsymbol{r} \cdot \boldsymbol{\sigma}$, where $\boldsymbol{r} = \begin{pmatrix} 1 & r_x & r_y & r_z \end{pmatrix}$ is a real-valued vector termed the *Bloch vector*. The factor of $\frac{1}{2}$ ensures that $\mathrm{Tr}(\rho) = 1$, and the positivity constraint $\rho \geq 0$ implies that $\sqrt{r_x^2 + r_y^2 + r_z^2} \leq 1$. Thus a general qubit state is fully specified by the coordinates (r_x, r_y, r_z), which must lie within a sphere of radius 1. The state space for a qubit is termed the *Bloch sphere*, displayed in Fig. 1.1(a). Pure states lie on the surface of the Bloch sphere ($\sqrt{r_x^2 + r_y^2 + r_z^2} = 1$), and mixed states are in the interior ($\sqrt{r_x^2 + r_y^2 + r_z^2} < 1$). The maximally-mixed state $\rho = \mathbb{I}/2$ has a Bloch vector with magnitude zero and it lies at the centre of the Bloch sphere.

Allowed unitary evolutions of a qubit have the form $\exp\left\{\frac{-i\hat{\mathbf{n}}\cdot\boldsymbol{\sigma}\phi}{2}\right\}$, which describes a rotation of the Bloch sphere by an angle ϕ about the axis defined by unit vector $\hat{\mathbf{n}}$ [6]. Note that such transformations cannot change the magnitude $|\boldsymbol{r}|$ of a Bloch vector, and hence unitary transformations cannot change the mixedness of a state.

A general measurement effect E can also be represented[2] as a linear combination of the identity and the Pauli matrices, $E = \boldsymbol{e} \cdot \boldsymbol{\sigma}$, where $\boldsymbol{e} = \begin{pmatrix} e_0 & e_x & e_y & e_z \end{pmatrix}$. The positivity constraint $\mathrm{Tr}\,\rho E \geq 0$ implies $e_0 \geq 0$ and $\sqrt{e_x^2 + e_y^2 + e_z^2} \leq e_0$, which defines a 4-dimensional cone with a spherical base that opens in the positive e_0 direction and has its vertex at

[2] Note that our definition of a qubit measurement effect $E = \mathbf{e} \cdot \boldsymbol{\sigma}$ differs from the standard convention used in quantum information theory, which is $E = \frac{1}{2}\mathbf{e} \cdot \boldsymbol{\sigma}$. Our choice of convention enables us to use the identity $\mathrm{Tr}(\rho E) = \boldsymbol{r} \cdot \boldsymbol{e}$, which will be useful in Chapters 3 and 5.

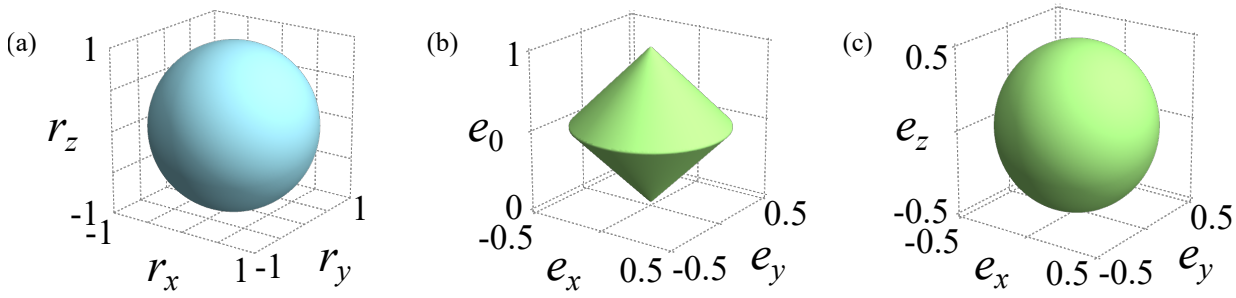

Figure 1.1: Bloch representation of qubit states and measurements. (a) The Bloch sphere represents the space of possible qubit states. (b), (c) Two three-dimensional projections of the four-dimensional qubit measurement effect space, which we refer to as the Bloch diamond.

$\boldsymbol{e} = (0,0,0,0)$. The constraint $\mathrm{Tr}(\rho E) \leq 1$ implies $e_0 \leq 1$ and $\sqrt{e_x^2 + e_y^2 + e_z^2} \leq 1 - e_0$, which is the 4-dimensional cone with spherical base that opens in the *negative* e_0 direction and has its vertex at $\boldsymbol{e} = (1,0,0,0)$. The allowed values of $\boldsymbol{e}$ lie in the intersection of these two cones, defining a 4-dimensional diamond with a spherical base, which we refer to as the *Bloch diamond* and display in Fig. 1.1(b,c). Rank-1 projective measurement effects (i.e. those which can be part of a rank-1 PVM), are found on the sphere defined by $e_0 = \frac{1}{2}$ and $\sqrt{e_x^2 + e_y^2 + e_z^2} = \frac{1}{2}$.

1.2 Quantization of the electromagnetic field

In this section we will present the quantum description of electromagnetic fields, and we follow the treatment presented in Refs. [7] and [8]. Classical electrodynamics is described by Maxwell's equations, which, in the absence of external fields or media (i.e. in vacuum), are:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{1.6}$$

$$\nabla \times \mathbf{B} = \mu_0 \varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \tag{1.7}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{1.8}$$

$$\nabla \cdot \mathbf{E} = 0, \tag{1.9}$$

where $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic field, and μ_0 and ε_0 are the vacuum permeability and permitivity, respectively.

Consider a single-mode electromagnetic wave resonant in a cavity of length L, propagating in the z-direction and polarized in the x-direction. The electric field of this wave,

$$\mathbf{E_x}(\mathbf{r},t) = \mathbf{e_x}\sqrt{\frac{2\omega^2}{V\varepsilon_0}}q(t)\sin(kz), \tag{1.10}$$

and corresponding magnetic field,

$$\mathbf{B_y}(\mathbf{r},t) = \mathbf{e_y}\frac{\mu_0\varepsilon_0}{k}\sqrt{\frac{2\omega^2}{V\varepsilon_0}}\frac{dq(t)}{dt}\cos(kz), \tag{1.11}$$

is a solution to Maxwell's equations. In the above, $\mathbf{e_x}$ ($\mathbf{e_y}$) is the unit vector along the x-axis (y-axis), and V is the effective volume of the cavity. ω is the frequency of the field, $c = \frac{1}{\sqrt{\mu_0\varepsilon_0}}$ is the speed of light in vacuum, and $k = \omega/c$ is the wavenumber of the field. We leave the quantity $q(t)$ undefined for now, but note that it has units of length.

To quantize the field we begin with its classical Hamiltonian, given by:

$$H = \frac{1}{2}\int dV\left[\varepsilon_0\mathbf{E}^2(\mathbf{r},t) + \frac{1}{\mu_0}\mathbf{B}^2(\mathbf{r},t)\right] \tag{1.12}$$

For our single-mode field, the Hamiltonian simplifies to:

$$H = \frac{1}{2}(p^2 + \omega^2 q^2), \tag{1.13}$$

where we have made the substitution $p(t) = \frac{dq(t)}{dt}$. The above Hamiltonian describes a simple harmonic oscillator, and we interpret $q(t)$ and $p(t)$ as generalized position and momentum, respectively. To quantize H, we make the replacement $q(t) \to \hat{q}(t)$ and $p(t) \to \hat{p}(t)$, which satisfy the canonical commutation relation $[\hat{q},\hat{p}] = i\hbar$. We now define the operators

$$\hat{a} = \frac{1}{\sqrt{2\hbar\omega}}(\omega\hat{q} + i\hat{p}), \tag{1.14}$$

$$\hat{a}^\dagger = \frac{1}{\sqrt{2\hbar\omega}}(\omega\hat{q} - i\hat{p}). \tag{1.15}$$

These have commutation relation $[\hat{a},\hat{a}^\dagger] = 1$ and can be used to rewrite Eq. (1.13) as

$$\hat{H} = \hbar\omega\left(\hat{a}^\dagger\hat{a} + \frac{1}{2}\right). \tag{1.16}$$

We can find the time-evolution of $\hat{a}$ using the relation

$$\frac{d\hat{a}}{dt} = \frac{-1}{i\hbar}[\hat{H}, \hat{a}], \tag{1.17}$$

which has the solution

$$\hat{a}(t) = \hat{a}(0)e^{-i\omega t}. \tag{1.18}$$

We can finally write the quantized form of the single-mode field (Eq. (1.10)) as:

$$\hat{\mathbf{E}}_{\mathbf{x}}(\mathbf{r}, t) = \mathbf{e}_{\mathbf{x}}\sqrt{\frac{\hbar\omega}{V\varepsilon_0}}(\hat{a}e^{-i\omega t} + \hat{a}^\dagger e^{i\omega t})\sin(kz), \tag{1.19}$$

where we have made the substitution $q \to \hat{q}$ and then the further substitution $\hat{q} = \sqrt{\frac{\hbar}{2\omega}}(\hat{a} + \hat{a}^\dagger)$.

A full multi-mode treatment of the field quantization, (and dropping the arguments $\mathbf{r}$ and t for notational simplicity), results in the following expression for the electric field [7]:

$$\mathbf{E} = i\sum_{\mathbf{k},s}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\varepsilon_0}}\mathbf{e}_{\mathbf{k},s}\left(\hat{a}_{\mathbf{k},s}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)} + \hat{a}^\dagger_{\mathbf{k},s}e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\right). \tag{1.20}$$

Here, each modes is labelled by its wavevector $\mathbf{k}$ and polarization s. There are two values for s, and they denote orthogonal polarizations. We will later find it convenient to write $\mathbf{E}$ as a sum of two components, $\mathbf{E} = \mathbf{E}^{(+)} + \mathbf{E}^{(-)}$, where $\mathbf{E}^{(+)} = i\sum_{\mathbf{k},s}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\varepsilon_0}}\mathbf{e}_{\mathbf{k},s}\hat{a}_{\mathbf{k},s}e^{i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$ and $\mathbf{E}^{(-)} = i\sum_{\mathbf{k},s}\sqrt{\frac{\hbar\omega_{\mathbf{k}}}{2\varepsilon_0}}\mathbf{e}_{\mathbf{k},s}\hat{a}^\dagger_{\mathbf{k},s}e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega t)}$, are the *positive* and *negative* frequency components of the electric field, respectively.

1.2.1 Number states

A number state (or Fock state) is an energy eigenstate of the quantized harmonic oscillator. We write $|n\rangle$ as the eigenstate of the simple harmonic oscillator Hamiltonian (Eq. (1.16)) with energy $\mathcal{E}_n$ such that $\hat{H}|n\rangle = \mathcal{E}_n|n\rangle$. Left-multiplying both sides of the eigenvalue equation by $\hat{a}$ and applying the commutation relation $[\hat{a}, \hat{a}^\dagger] = 1$ gives:

$$\hat{H}\hat{a}|\psi_n\rangle = (\mathcal{E}_n - \hbar\omega)a|n\rangle. \tag{1.21}$$

A similar equation can be found by mutiplying by $\hat{a}^\dagger$ instead:

$$\hat{H}\hat{a}^\dagger|\psi_n\rangle = (\mathcal{E}_n + \hbar\omega)\hat{a}^\dagger|n\rangle. \tag{1.22}$$

Thus, $\hat{a}^\dagger$ adds an excitation of energy $\hbar\omega$ to the field, and $\hat{a}$ removes an excitation of the same energy. We interpret an excitation of energy $\hbar\omega$ as a *photon*, and we call $\hat{a}$ and $\hat{a}^\dagger$ the *annihilation* and *creation* operator, respectively.

We denote $|n-1\rangle$ and $|n+1\rangle$ as the eigenstates with energies $\mathcal{E}_n - \hbar\omega$ and $\mathcal{E}_n + \hbar\omega$, respectively. The specific effect $\hat{a}$ and $\hat{a}^\dagger$ have on energy eigenstate $|n\rangle$ are [7]:

$$\hat{a}\,|n\rangle = \sqrt{n}\,|n-1\rangle \tag{1.23}$$

$$\hat{a}^\dagger\,|n\rangle = \sqrt{n+1}\,|n+1\rangle\,. \tag{1.24}$$

This implies that $\langle n|\,\hat{a}^\dagger\hat{a}\,|n\rangle = n$, and we define the number operator $\hat{n} = \hat{a}^\dagger a$. The ground (or *vacuum*) state of the quantum harmonic oscillator, denoted $|0\rangle$, has energy $\frac{\hbar\omega}{2}$. Eigenstate $|n\rangle$ is a state of n photons, and it has energy $\hbar\omega(n+\frac{1}{2})$.

1.2.2 Coherent states

Another interesting state of light is the coherent state, which is a good representation of the light emitted by a laser [9]. A coherent state $|\alpha\rangle$ is defined as an eigenstate of the annihilation operator $\hat{a}$, with eigenvalue α. That is, $\hat{a}\,|\alpha\rangle = \alpha\,|\alpha\rangle$. We can write a coherent state as a superposition of Fock states:

$$|\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2}\sum_{n=0}^{\infty}\frac{\alpha^n}{\sqrt{n!}}\,|n\rangle\,. \tag{1.25}$$

Here, α is a complex number that represents the amplitude of the coherent state. The mean number of photons in a coherent state is given by the expectation value of the number operator, $\langle\alpha|\,\hat{n}\,|\alpha\rangle = \langle\alpha|\,\hat{a}^\dagger\hat{a}\,|\alpha\rangle = |\alpha|^2$. The variance in the photon number is given by $\langle\alpha|\,\hat{n}^2\,|\alpha\rangle - \langle\alpha|\,\hat{n}\,|\alpha\rangle^2 = |\alpha|^2$, which is equal to the mean photon number. In fact, the photon number distribution of a coherent state is represented by a Poisson distribution with mean $|\alpha|^2$.

Consider the operator $\hat{a}\hat{a}^\dagger$. Its expectation value for coherent state $|\alpha\rangle$ is $\langle\alpha|\,\hat{a}\hat{a}^\dagger\,|\alpha\rangle = \langle\alpha|\,\hat{a}^\dagger\hat{a}+1\,|\alpha\rangle = |\alpha|^2+1$. In the *strong coherent state limit* $|\alpha|^2 \gg 1$, we can make the approximation $|\alpha|^2+1 \approx |\alpha|^2$, or equivalently, $\hat{a}^\dagger\,|\alpha\rangle \approx \alpha^*\,|\alpha\rangle$. We will make use of this approximation in Section 1.3.1.

1.2.3 Second-order coherence measurement

Consider the number state $|1\rangle$ and the coherent state $|\alpha\rangle$ with $|\alpha|^2 = 1$. Their overlap is $|\,\langle 1\mid\alpha\rangle\,|^2 = e^{-1} \approx 0.368$. However, a measurement of the average photon number

cannot distinguish these states, since they both have mean photon number $\langle \hat{n} \rangle = 1$. Thus some other measurement is required in order to tell these two states apart. One such measurement [8, 10] is the second-order coherence, $g^{(2)}(0)$, which is defined as[3]:

$$g^{(2)}(0) = \frac{\langle \hat{a}^\dagger \hat{a}^\dagger \hat{a} \hat{a} \rangle}{\langle \hat{a}^\dagger \hat{a} \rangle \langle \hat{a}^\dagger \hat{a} \rangle}. \tag{1.26}$$

A simple calculation shows that for the coherent state $|\alpha\rangle$ we have $g^{(2)}(0) = 1$, regardless of the value of α, and for a number state $|n\rangle$ we have $g^{(2)}(0) = \frac{n(n-1)}{n^2} = 1 - \frac{1}{n}$. Hence our coherent state with $\alpha = 1$ has a second-order correlation of 1, and our single-photon state $|1\rangle$ has a second-order correlation of 0. Measuring the second-order coherence function of a photon source can be an effective way to characterize the single-photon nature of the source.

Hanbury-Brown–Twiss interferometer

In order to measure the second-order coherence we use the Hanbury-Brown–Twiss interferometer [10] pictured in Fig. 1.2(a). The source emits light into the signal mode, s. The signal mode enters one of two inputs to a beamsplitter, with vacuum in the other input mode, v. After propagating through the beamsplitter, the input modes are transformed into the output modes $\hat{a}_1$ and $\hat{a}_2$ according to:

$$\hat{a}_s \to \frac{1}{\sqrt{2}}(\hat{a}_1 + \hat{a}_2) \tag{1.27}$$

$$\hat{a}_v \to \frac{1}{\sqrt{2}}(\hat{a}_1 - \hat{a}_2). \tag{1.28}$$

We can also define the corresponding *back-propagating* transformations:

$$\hat{a}_1 \to \frac{1}{\sqrt{2}}(\hat{a}_s + \hat{a}_v) \tag{1.29}$$

$$\hat{a}_2 \to \frac{1}{\sqrt{2}}(\hat{a}_s - \hat{a}_v). \tag{1.30}$$

[3] One can also define a second-order coherence that depends on a time delay τ as $g^{(2)}(\tau) = \frac{\langle \hat{a}^\dagger(t)\hat{a}^\dagger(t+\tau)\hat{a}(t)\hat{a}(t+\tau)\rangle}{\langle \hat{a}^\dagger(t)\hat{a}(t)\rangle\langle \hat{a}^\dagger(t+\tau)\hat{a}(t+\tau)\rangle}$ [8, 10]. In this thesis we will only be concerned with time delays of $\tau = 0$, and we drop the t dependence on the right-hand side of the equation to simplify the notation. However, we write $g^{(2)}(0)$ to be clear that we are considering the case with zero time delay.

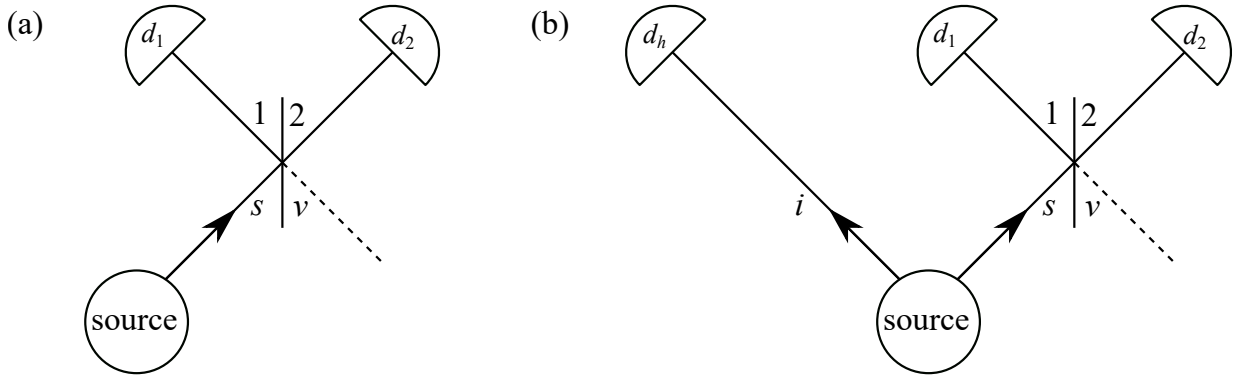

Figure 1.2: Hanbury-Brown–Twiss interferometer for measurement of the second-order coherence function. (a) Setup for measurement of second-order coherence, $g^{(2)}(0)$. (b) Setup for measurement of heralded second-order coherence, $g_h^{(2)}(0)$.

The detectors used are *bucket detectors*, which "click" with some probability η when one or more photons are incident upon them, and otherwise they "don't click"[4]. Such a detector can be modelled by the POVM $\{\eta(\mathbb{I} - |0\rangle\langle 0|), |0\rangle\langle 0| + (1-\eta)(\mathbb{I} - |0\rangle\langle 0|)\}$. Note that back-propagation through the beamsplitter transforms the quantity $\frac{\langle \hat{n}_{12}\rangle}{\langle \hat{n}_1\rangle\langle \hat{n}_2\rangle}$ as:

$$\frac{\langle \hat{n}_{12}\rangle}{\langle \hat{n}_1\rangle \langle \hat{n}_2\rangle} = \frac{\left\langle \hat{a}_1^\dagger \hat{a}_2^\dagger \hat{a}_1 \hat{a}_2 \right\rangle}{\left\langle \hat{a}_1^\dagger \hat{a}_1 \right\rangle \left\langle \hat{a}_2^\dagger \hat{a}_2 \right\rangle} \tag{1.31}$$

$$= \frac{\frac{1}{4}\left\langle (\hat{a}_s^\dagger + \hat{a}_v^\dagger)(\hat{a}_s^\dagger - \hat{a}_v^\dagger)(\hat{a}_s + \hat{a}_v)(\hat{a}_s - \hat{a}_v) \right\rangle}{\frac{1}{2}\left\langle (\hat{a}_s^\dagger + \hat{a}_v^\dagger)(\hat{a}_s + \hat{a}_v) \right\rangle \frac{1}{2}\left\langle (\hat{a}_s^\dagger - \hat{a}_v^\dagger)(\hat{a}_s - \hat{a}_v) \right\rangle} \tag{1.32}$$

$$= \frac{\left\langle \hat{a}_s^\dagger \hat{a}_s^\dagger \hat{a}_s \hat{a}_s \right\rangle}{\left\langle \hat{a}_s^\dagger \hat{a}_s \right\rangle \left\langle \hat{a}_s^\dagger \hat{a}_s \right\rangle}, \tag{1.33}$$

which has the same form as Eq. (1.26). To go from the second to third line above we used the fact that input mode v is in the vacuum state. If we measure detector clicks for a short period of time w (which we call a *coincidence window*), then we can interpret $\frac{\langle \hat{n}_{12}\rangle}{\langle \hat{n}_1\rangle\langle \hat{n}_2\rangle}$ as $\frac{p(12)}{p(1)p(2)}$, where $p(1)$ ($p(2)$) is the probability of detector d_1 (d_2) emitting a click during the window, and $p(12)$ is the probability of both detectors clicking during the

[4]In practice, all detectors also have *dark counts*, which are false "clicks" that can be produced even when no light is incident upon the detectors. Here we will assume that these dark counts are negligible.

window. These probabilities can be measured[5] by recording the number of single (n_1 and n_2) and coincident (n_{12}) clicks for some set period of time t and dividing by the number of coincidence windows that occurred during the measurement time (t/w), such that:

$$g^{(2)}(0) = \frac{p(12)}{p(1)p(2)} = \frac{n_{12}/(\frac{t}{w})}{(n_1/(\frac{t}{w}))(n_2/(\frac{t}{w}))} = \frac{n_{12}(\frac{t}{w})}{n_1 n_2}. \tag{1.34}$$

Some light sources are operated in a *heralded* manner. Such sources will emit light in two modes, one labelled the signal mode, and the other the *idler* mode, and some event in the idler mode will herald the presence of the desired state in the signal mode. The second-order correlation of such sources must also be measured in a heralded manner, and is defined as [11, 12]:

$$g_h^{(2)}(0) = \frac{p(12|h)}{p(1|h)p(2|h)}, \tag{1.35}$$

where we have now conditioned the probabilities $p(12)$, $p(1)$, and $p(2)$ on the occurence of the heralding event. These probabilities can be expressed as $p(12|h) = n_{12h}/n_h$, $p(1|h) = n_{1h}/n_h$, and $p(2|h) = n_{2h}/n_h$, where n_h represents the number of heralding events, and the n's with multiple subscripts represent coincident detection and heralding events. This leads us to the following expression for $g_h^{(2)}(0)$, in terms of experimentally observed events:

$$g_h^{(2)}(0) = \frac{n_{12h} n_h}{n_{1h} n_{2h}}. \tag{1.36}$$

The setup in Fig. 1.2(b) can be used to measure the heralded second-order coherence function, if we define a "click" from detector d_h as the heralding event.

1.3 Three-wave mixing

One method of creating photon number states is to impinge a coherent laser beam on a nonlinear crystal where it will be probabilistically transformed into photon pairs via the process called *spontaneous parametric downconversion* (SPDC). This process is described by the *three-wave mixing* Hamiltonian. We will very briefly discuss the classical description of three-wave mixing, and then discuss its quantum description.

When a classical electric field $\mathbf{E}(t)$ interacts with a nonlinear medium it can induce a polarization $\mathbf{P}(t)$ in the medium that depends on the input field in a nonlinear way. We

[5]The described method for measuring $p(12)$, $p(1)$, and $p(2)$ is valid for a weak source with $p(1)$ and $p(2) \ll 1$.

can expand the i-th component of $\mathbf{P}(t)$, $P_i(t)$, as a function of powers of the components of $\mathbf{E}(t)$ [13] as follows:

$$P_i(t) = \varepsilon_0(\chi^{(1)}_{ij}E_j(t) + \chi^{(2)}_{ijk}E_j(t)E_k(t) + \chi^{(3)}_{ijkl}E_j(t)E_k(t)E_l(t) + \dots), \tag{1.37}$$

where summation over repeated indices is implied, and the coefficients $\chi^{(i)}$ are tensors that determine the strength of each term of the expansion.

The classical energy density, U, of the electric field, is given by [13]:

$$U = \frac{1}{2}\left\langle(\varepsilon_0\mathbf{E} + \mathbf{P})\cdot\mathbf{E}\right\rangle_t, \tag{1.38}$$

where $\langle\cdot\rangle_t$ denotes a time average. If the input electric field has multiple frequency components such that $E_i(t) = \sum_n E_i(\omega_n)e^{-i\omega t}$, then the expression for U becomes [13]:

$$U = \frac{\varepsilon_0}{2}\sum_n \chi^{(1)}_{ij}E^*_i(\omega_n)E_j(\omega_n) + \frac{\varepsilon_0}{3}\sum_{mn}\chi^{(2)}_{ijk}E^*_i(\omega_m+\omega_n)E_j(\omega_m)E_k(\omega_n) + \cdots, \tag{1.39}$$

where we again sum over repeated indices, and have assumed that the $\chi^{(i)}$'s have negligible frequency dependence. The $\chi^{(1)}$ term describes conventional linear optics. The next term describes mixing between fields with frequencies ω_m, ω_n, and $\omega_m + \omega_n$, which is referred to as *three-wave mixing*. The classical hamiltonian H is found by integrating the energy density over the effective volume of the fields, $H = \int_V dVU$.

We will consider the case where a pump beam with frequency ω_p interacts with beams in two lower-frequency modes ω_s and ω_i, such that $\omega_p = \omega_s + \omega_i$. Furthermore, we will assume that each field is polarized so we can replace the term $\chi^{(2)}_{ijk}E^*_i(\omega_p)E_j(\omega_s)E_k(\omega_i)$ with $\chi^{(2)}_{eff}E^*_pE_sE_i$, where the subscripts p, s, and i now label the three different modes. If we separate H into the linear term H_0 and the nonlinear term H', such that $H = H_0 + H'$, and truncate our expansion at the $\chi^{(2)}$ term, then we have:

$$H' = \frac{\varepsilon_0}{3}\chi^{(2)}_{eff}\int_V dV\chi^{(2)}_{eff}E^*_pE_sE_i. \tag{1.40}$$

1.3.1 Spontaneous parametric downconversion

The quantized three-wave mixing Hamiltonian $\hat{H}$ is found by moving to the interaction picture [8] and making the substitution $E \to \hat{E}^{(+)} + E^{(-)}$ in Eq. (1.40), to arrive at:

$$\hat{H}' = \frac{\varepsilon_0}{3}\chi^{(2)}_{eff}\int_V dV\hat{E}^{(-)}_s\hat{E}^{(-)}_i\hat{E}^{(+)}_p + \hat{E}^{(+)}_s\hat{E}^{(+)}_i\hat{E}^{(-)}_p, \tag{1.41}$$

where we have only kept the two terms which conserve energy.

We will assume that the signal and idler modes are initially in the vacuum state, and that the pump mode contains a strong coherent laser beam $|\alpha\rangle_p$ with $|\alpha|^2 \gg 1$. We will assume the interaction is weak, and that α stays relatively constant throughout the interaction (this is called the *undepleted pump approximation*). We use the strong coherent state approximation to make the substitution $\hat{a}_p |\alpha\rangle_p \to \alpha^* |\alpha\rangle_p$ (see Section 1.2.2). Finally, remembering that $E^{(+)} \propto \hat{a}$ and $E^{(-)} \propto \hat{a}^\dagger$, we can rewrite $\hat{H}'$ as:

$$\hat{H}' = g\hat{a}_s^\dagger \hat{a}_i^\dagger + g^* \hat{a}_s \hat{a}_i, \tag{1.42}$$

where we have absorbed all constants above into the single complex number g. We note that $g \propto \alpha*$, and thus the strength of the interaction can be tuned with the strength of the pump field.

Thus the initial state $|\psi(0)\rangle = |0\rangle_s |0\rangle_i$ evolves under $\hat{H}'$ as

$$|\psi(t)\rangle = \exp\left\{-\frac{i}{\hbar} t(g\hat{a}_s^\dagger \hat{a}_i^\dagger + g^* \hat{a}_s \hat{a}_i)\right\} |0\rangle_s |0\rangle_i . \tag{1.43}$$

This evolution is identical to the two-mode squeezing unitary $U(\xi) = \exp\left\{\xi \hat{a}_s^\dagger \hat{a}_i^\dagger - \xi^* \hat{a}_s \hat{a}_i)\right\}$, if we make the substitution $-\frac{i}{\hbar} tg = re^{i\theta} = \xi$. The two-mode squeezing operator has the following effect on the initial state: [7]

$$U(\xi) |0\rangle_s |0\rangle_i = \sqrt{1 - |\gamma|^2} \left(|0\rangle_s |0\rangle_i - \gamma |1\rangle_s |1\rangle_i + \gamma^2 |2\rangle_s |2\rangle_i - \dots\right), \tag{1.44}$$

where $\gamma = -e^{i\theta} \tanh r$. The interpretation of the $\chi^{(2)}$ interaction is that, with some probability, a single photon in the pump beam is downconverted into a pair of signal and idler photons. Since this process is probabilistic, the probability of downconverting two pairs is the square of the probability of downconverting one pair. For a weak interaction, the signal and idler modes after the crystal are mostly vacuum, and the next leading-order term is a single photon pair.

1.3.2 Heralded SPDC source

If we place an ideal detector in the idler arm, wait for it to "click", and then trace out the idler mode, the state remaining in the signal mode will be:

$$\rho_s \propto \mathrm{Tr}_i[(\mathbb{I} - |0\rangle \langle 0|) U(\xi) |0\rangle_s |0\rangle_i] \tag{1.45}$$

$$= (1 - |\gamma|^2) \left(|1\rangle_s \langle 1|_s + |\gamma|^2 |2\rangle_s \langle 2|_s + \dots\right), \tag{1.46}$$

which has a high overlap with the single photon state $|1\rangle_s$ when $|\gamma|^2$ is not too large. Thus, heralding an SPDC state is an effective way of producing a single photon state.[6]

We can estimate the relative frequency of single to double pairs in the SPDC state with the heralded second-order coherence measurement (see Fig. 1.2(b)) and Eq. (1.36)). We assume that $|\gamma| \ll 1$, and neglect the terms with three or more photon pairs. We also assume that we use a continuous wave pump laser, which implies that the downconversion probability is constant in time[7]. Finally, we choose a coincidence window length w that is much greater than the coherence time of the signal and idler mode photons. Under this condition, we can assume that if two pairs are created during one coincidence window they will not interfere with each other. The efficiencies of detectors d_1, d_2, and d_h are η_1, η_2, and η_h, respectively. Let $\bullet$ denote the case when the source produces one pair, and $\bullet\bullet$ the case when two pairs are produced. These events have probabilities $p(\bullet) \propto |\gamma|^2$ and $p(\bullet\bullet) \propto |\gamma|^4$. The expected numbers of measured counts are proportional to the following probabilities:

$$n_h \propto p(h|\bullet)p(\bullet) + p(h|\bullet\bullet)p(\bullet\bullet) \tag{1.47}$$

$$n_{1h} \propto p(1|\bullet)p(h|\bullet)p(\bullet) + p(1|\bullet\bullet)p(h|\bullet\bullet)p(\bullet\bullet) \tag{1.48}$$

$$n_{2h} \propto p(2|\bullet)p(h|\bullet)p(\bullet) + p(2|\bullet\bullet)p(h|\bullet\bullet)p(\bullet\bullet) \tag{1.49}$$

$$n_{12h} \propto p(12|\bullet)p(h|\bullet)p(\bullet) + p(12|\bullet\bullet)p(h|\bullet\bullet)p(\bullet\bullet). \tag{1.50}$$

The probability of detector d_h clicking when one pair is produced is $p(h|\bullet) = \eta_h$. For detectors d_1 and d_2 the corresponding probabilities are $p(i|\bullet) = \frac{\eta_i}{2}$ (for $i = 1, 2$), where the extra factor of $\frac{1}{2}$ is due to the 50% transmission/reflection probability at the beamsplitter. When two pairs are produced, the probability of detecting a coincidence between detectors d_1 and d_2 is the probability that the signal photon from the first pair is transmitted at the beamsplitter and detected by d_1 multiplied by the probability that the photon from the second pair is detected at d_2, *plus* the probability that the first signal photon is detected by d_2 times the probability that the second is detected by d_1. Thus, $p(12|\bullet\bullet) = \frac{\eta_1}{2}\frac{\eta_2}{2} + \frac{\eta_2}{2}\frac{\eta_1}{2}$. Finally, $p(h|\bullet\bullet)$, is the probability that the idler photon from the first pair makes d_h click, plus the probability that the idler photon from the first pair *doesn't* make d_h click multiplied by the probability that the idler photon from the second pair does [16]. Thus, $p(h|\bullet\bullet) = \eta_h + (1 - \eta_h)\eta_h \approx 2\eta_h$, where we have made the additional approximation that $\eta_h \ll 1$.

Substituting these probabilities into Eqs. (1.47)-(1.50), and only keeping terms of

[6]The specific heralded SPDC source used for the experiments in this thesis is detailed in Refs. [14] and [15].

[7]If the pump laser is pulsed, then photon pairs are only produced during the lifetime of the pump pulses, and not in between pulses.

lowest-order in $|\gamma|^2$ we find:

$$n_h \propto |\gamma|^2 \eta_h \tag{1.51}$$

$$n_{1h} \propto |\gamma|^2 \frac{\eta_1}{2} \eta_h \tag{1.52}$$

$$n_{2h} \propto |\gamma|^2 \frac{\eta_2}{2} \eta_h \tag{1.53}$$

$$n_{12h} \propto |\gamma|^4 \eta_1 \eta_2 \eta_h, \tag{1.54}$$

$$\tag{1.55}$$

which implies that, for a heralded SPDC source:

$$g_h^{(2)}(0) \approx 4|\gamma|^2. \tag{1.56}$$

Finally, we arrive at the ratio of single pairs to double pairs produced by the source, $R_{1:2}$, as a function of the heralded $g^{(2)}$ measurement:

$$R_{1:2} \approx \frac{1}{|\gamma|^2} \approx \frac{4}{g_h^{(2)}(0)}. \tag{1.57}$$

1.4 Photon polarization

A photon's polarization state is the direction in which the electric field oscillates, and it is possible to manipulate polarization to encode quantum information. If the electric field oscillates horizontally (with respect to some reference, say for example the plane of the optical table over which the photon is travelling) we can say the photon is in polarization state $|H\rangle$. Conversely, if the field instead oscillates vertically, we say the photon has polarization state $|V\rangle$. A photon can also be in a superposition of polarization states. For example, diagonally polarized light has state $|D\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)$, and anti-diagonally polarized light has state $|A\rangle = \frac{1}{\sqrt{2}}(|H\rangle - |V\rangle)$. We can also take superpositions with complex coefficients: right-circular polarized light has state $|R\rangle = \frac{1}{\sqrt{2}}(|H\rangle + i\,|V\rangle)$, and left-circular light has state $|L\rangle = \frac{1}{\sqrt{2}}(|H\rangle - i\,|V\rangle)$.

The states $|H\rangle$ and $|V\rangle$ are orthogonal, and if we make the assignments $|H\rangle = |0\rangle$ and $|V\rangle = |1\rangle$, we see that $|D\rangle = |+\rangle$, $|A\rangle = |-\rangle$, $|R\rangle = |+i\rangle$, and $|L\rangle = |-i\rangle$. The full space of allowed polarization states forms a sphere, called the Poincaré sphere [17]. There is a one-to-one correspondance between the Bloch sphere and the Poincaré sphere, and thus a polarization-encoded single photon is a qubit.

1.4.1 Manipulating polarization with birefringent materials

Polarization states can be manipulated with birefringent materials, that is, materials for which orthogonal polarizations travel at different speeds. A *waveplate* is a birefringent material, which has fast and slow axes that are perpendicular to each other. Light polarized parallel to the fast axis experiences one index of refraction n_f while it is in the waveplate, and light polarized parallel to the slow axis experiences index n_s. If the waveplate is aligned such that the fast axis is aligned horizontally, then the vertically-polarised component of a photon passing through it is delayed relative to the horizontal component by a phase of $\phi = \frac{n_s\omega}{c}L - \frac{n_f\omega}{c}L$. This transformation is captured by the unitary,

$$U_\phi = \begin{pmatrix} e^{i\frac{n_f\omega}{c}L} & 0 \\ 0 & e^{i\frac{n_s\omega}{c}L} \end{pmatrix}, \tag{1.58}$$

$$= \begin{pmatrix} 1 & 0 \\ 0 & e^{-i\phi} \end{pmatrix}, \tag{1.59}$$

where we have ignored an unimportant global phase to write the second line. If the waveplate is rotated by an angle θ about the direction of light propagation then the unitary becomes:

$$U_\phi(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & e^{-i\phi} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \tag{1.60}$$

$$= \begin{pmatrix} \cos^2\theta + e^{-i\phi}\sin^2\theta & \cos\theta\sin\theta(1-e^{-i\phi}) \\ \cos\theta\sin\theta(1-e^{-i\phi}) & e^{-i\phi}\cos^2\theta + \sin^2\theta \end{pmatrix} \tag{1.61}$$

A waveplate for which $\phi = \pi$ is called a *half waveplate*, with unitary:

$$U_H(\theta) = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}. \tag{1.62}$$

Similary, a *quarter waveplate* ($\phi = \pi/2$) has unitary:

$$U_Q(\theta) = \begin{pmatrix} \cos^2\theta - i\sin^2\theta & (1+i)\cos\theta\sin\theta \\ (1+i)\cos\theta\sin\theta & -i\cos^2\theta + \sin^2\theta \end{pmatrix}, \tag{1.63}$$

A general single-qubit unitary U can be decomposed as a product of three waveplate unitaries, $U = U_Q(\theta_3)U_H(\theta_2)U_Q(\theta_1)$, and thus an arbitrary polarization rotation can be perfomed with one half and two quarter waveplates [18]. The related problem of transforming

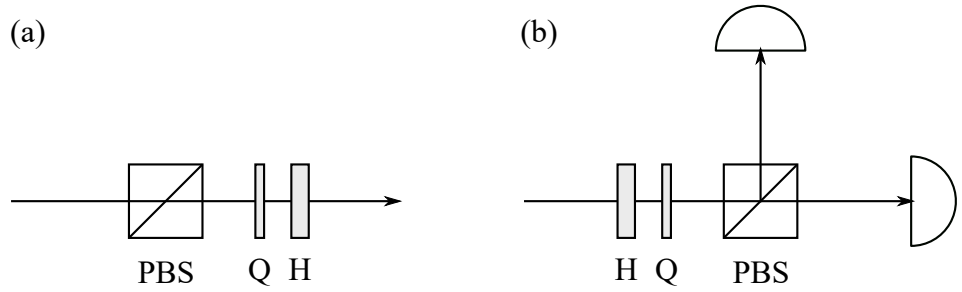

Figure 1.3: Preparation and measurement of polarization states of light. (a) A polarizing beamsplitter, quarter waveplate, and half waveplate can prepare any pure polarization state. (b) A half waveplate, quarter waveplate, polarizing beamsplitter, and two detectors can measure any polarization state. PBS: polarizing beamsplitter; Q: quarter waveplate; H: half waveplate.

a known linear polarization state into an arbitrary pure target state can be solved using only two waveplates [19, 17]. For example, if we start with horizontally polarized light and pass it through a quarter waveplate and then a half waveplate, any pure polarization state can be created by independently tuning the waveplate angles. This process also works in reverse: an arbitrary pure polarization state can be rotated to horizontal polarization with a half and then quarter waveplate. With the addition of a polarizing beamsplitter (a device which transmits horizontally-polarized light and reflects vertically-polarized light), we can prepare and measure any polarization state, with a setup as depicted in Fig. 1.3.

1.5 Maximum likelihood quantum tomography

The process of inferring, from data, a quantum description of an experimental procedure is called *quantum tomography*. Quantum tomography can be divided into three main categories: state [20, 21, 22, 23, 24], process [25, 26, 27, 28], and measurement [29, 30, 31, 32] tomography.

In this section we present tomography methods that rely on the *maximum likelihood principle*. This principle states that the best estimate of the set of operators describing a set of data is the set of operators that has the highest likelihood of producing that data. Thus, for maximum likelihood quantum tomography, the tomography procedure always contains an optimization step for which one has to find an optimal set of quantum states, transformations, and/or measurements among the state, transformation and measurement spaces allowed by quantum theory.

In this section we will describe the minimal sets of data that must be collected for the different types of maximum likelihood quantum tomography, and then define the optimization problem that must be solved.

1.5.1 Quantum state tomography

A d-level quantum system is described by a $d \times d$ density operator ρ. The constraint $\mathrm{Tr}\,\rho = 1$ implies that only $d-1$ real numbers are necessary to uniquely determine the d diagonal entries, and $d(d-1)$ real numbers are needed for the $d(d-1)/2$ complex numbers above the diagonal. Since ρ is Hermitian, the entries above the diagonal determine the entries below the diagonal. Thus ρ is described by d^2-1 real parameters.

A set of d^2-1 linearly-independent measurement effects is called *tomographically complete* for a d-level quantum system, because the measurement-outcome probabilities for such a set of effects uniquely determine ρ. (In practice, d^2 effects are often required for normalization). These measurement-outcome probabilities can only be approximately measured in practice, due to statistical noise in the measurements.

Consider an experiment in which many copies of an unknown state ρ are created, and each state is measured by one of m measurement devices M_i, for $i \in \{1, \ldots, m\}$. The i-th measurement M_i has K outcomes, and it is described by the (known) set of POVM effects $\{E_1^{(i)}, \ldots, E_K^{(i)}\}$. The set of measurements M_i is tomographically complete if the set of all effects describing all measurements M_i contains a set of d^2 linearly independent effects. Assume N copies of the state are measured by device M_i, and let $n_k^{(i)}$ be the number of times outcome k was returned. Then, if $\vec{n}$ is a string describing the measurement outcomes for *all* measurements on all copies of the state, the probability of obtaining $\vec{n}$ is [21]:

$$p(\vec{n}|\rho) = \prod_{i=1}^{m} \prod_{k=1}^{K} \mathrm{Tr}(\rho E_k^{(i)})^{n_k^{(i)}}, \tag{1.64}$$

where ρ is the *unknown* quantum state of the system.

The maximum likelihood estimate $\hat{\rho}$ of the true density matrix ρ, given the data $\vec{n}$, is the operator $\hat{\rho}$ which maximizes the quantity $p(\vec{n}|\hat{\rho})$. Thus, the procedure for maximum likelihood quantum state tomography is to first gather measurement-outcome statistics for d^2 linearly independent measurement effects, then find the $\hat{\rho}$ maximizing Eq. (1.64). In practice, to ensure that $\hat{\rho}$ is positive-definite, Hermitian, and has trace 1, it can be parameterized using the Cholesky decomposition $\hat{\rho} = \frac{T^\dagger T}{\mathrm{Tr}(T^\dagger T)}$, where T is a $d \times d$ lower-triangular complex matrix with real numbers on the diagonal [21].

Single qubit maximum likelihood state tomography

If we assume an explicit error model for an experiment then Eq. (1.64) can be simplified. Let's consider the example of measuring the polarization of a single photon [24]. We will assume that we have a single photon source that produces photons in the (unknown) polarization state ρ with some constant average rate, and that we can send these photons into a measurement device and count the number of photons for which a specific outcome k is returned. The number of times, n_k, that the k-th outcome is observed is an estimate of the expected number of outcomes $\bar{n}_k = N \operatorname{Tr}(\rho E_i)$. N can be estimated by measuring the source's photon production rate, the overall system efficiency, and the measurement time.

The number statistics of photons produced via SPDC in a nonlinear crystal pumped by a strong coherent state are described by a Poissonian distribution. We will assume that the expected number of outcomes $\bar{n}_k$ is large enough to approximate the distribution of observed outcomes n_k by the Gaussian distribution with mean and variance $\bar{n}_k$:

$$p(n_k|\rho) = \frac{1}{N_{\text{norm}}} \exp\left\{-\frac{(n_k - \bar{n}_k)^2}{2\bar{n}_k}\right\} \tag{1.65}$$

where N_{norm} is a normalization constant.

Now consider performing a series of *independent* measurements of the probabilities $\operatorname{Tr}(\rho E_k)$ for a tomographically complete set of measurement effects $\{E_1, \ldots, E_n\}$. (Here, the set $\{E_1, \ldots, E_n\}$ doesn't have to represent the set of effects for a single POVM, it can be a set of effects from multiple POVMs.) If $\vec{n}$ is a string encoding the observed measurement outcomes, the likelihood function is given by:

$$p(\vec{n}|\rho) = \frac{1}{N_{\text{norm}}} \prod_{k=1}^{n} \exp\left\{-\frac{(n_k - \bar{n}_k)^2}{2\bar{n}_k}\right\} \tag{1.66}$$

If the photon production rate and overall system efficiency stay constant throughout the data-collection process, this function becomes:

$$p(\vec{n}|\rho) = \frac{1}{N_{\text{norm}}} \prod_{k=1}^{n} \exp\left\{-\frac{(n_k - N \operatorname{Tr}(\rho E_k))^2}{2N \operatorname{Tr}(\rho E_k)}\right\}. \tag{1.67}$$

The operator $\hat{\rho}$ that maximizes the above function is the maximum likelihood estimate of the true state ρ. The problem of maximizing the above likelihood function is equivalent to

maximizing its logarithm, or equivalently, solving the following minimization problem [24]:

$$\underset{\hat{\rho}}{\text{minimize}} : \sum_{k=1}^{n} \frac{(n_k - N \operatorname{Tr}(\hat{\rho} E_k))^2}{N \operatorname{Tr}(\hat{\rho} E_k)}. \tag{1.68}$$

In practice, the minimization is solved by numerical methods, and the parameterization $\hat{\rho} = \frac{T^\dagger T}{\operatorname{Tr}(T^\dagger T)}$ is used to ensure that the positivity and trace-one constraints on ρ are met.

1.5.2 Quantum measurement tomography

Measurement tomography is the dual problem to state tomography. The goal is to infer the quantum description of an unknown measurement effect from data collected by measuring a known set of states.

Consider the problem of characterizing an unknown measurement device in a d-dimensional Hilbert space with K outcomes. This device is described by a POVM with K effects, $\{E_1, \ldots, E_K\}$. Each effect is a $d \times d$, positive, Hermitian matrix, that is described by d^2 real parameters, and hence a tomographically complete set of states contains d^2 linearly independent members. The set of measurement effects have the additional constraint that they must sum to the identity (i.e. $\sum_{k=1}^{K} E_k = \mathbb{I}$).

Let $\rho^{(i)}$ be a member of a tomographically complete set of states $\{\rho^{(1)}, \ldots, \rho^{(m)}\}$. Assume that many copies of a system prepared in state $\rho^{(i)}$ are sent into the uncharacterized measurement device, and that each outcome is observed $n_k^{(i)}$ times. If $\vec{n}$ is a string that encodes the full set of observed outcomes, then the probability of obtaining a specific value of $\vec{n}$ is [30]:

$$p(\vec{n}|\{E_1, \ldots, E_K\}) = \prod_{i=1}^{m} \prod_{k=1}^{K} \operatorname{Tr}(\rho^{(i)} E_k)^{n_k^{(i)}}. \tag{1.69}$$

The maximum likelihood estimate of the true POVM is the set $\{\hat{E}_1, \ldots, \hat{E}_K\}$ that maximizes $p(\vec{n}|\{\hat{E}_1, \ldots, \hat{E}_K\})$.

To ensure that each measurement effect is positive and Hermitian, we once again use the parameterization $E_k = T_k^\dagger T_k$ for the first $K-1$ effects (where, again, T_k is a lower-triangular complex matrix with real entries on the diagonal). The K-th effect is determined by the first $K-1$ effects, and is given by:

$$E_K = \mathbb{I} - \sum_{k=1}^{K-1} T_k^\dagger T_k. \tag{1.70}$$

This parameterization doesn't guarantee that the effect E_K will be positive, so the constraint $E_K \geq 0$ must be included when maximizing Eq. (1.69).

If we assume Poissonian counting statistics (as we did above for state tomography), we can recast the maximum-likelihood quantum measurement tomography problem as the following minimization problem:

$$\underset{\{\hat{E}_1,\ldots,\hat{E}_K\}}{\text{minimize}} : \sum_{i=1}^{m}\sum_{k=1}^{K} \frac{(n_k^{(i)} - N\,\mathrm{Tr}\,(\rho^{(i)}\hat{E}_k))^2}{N\,\mathrm{Tr}\,(\rho^{(i)}\hat{E}_k)}, \tag{1.71}$$

$$\text{subject to} : E_K \geq 0. \tag{1.72}$$

1.5.3 Quantum process tomography

The final piece in the quantum tomography puzzle is quantum process tomography, which is the process of characterizing an unknown quantum transformation from a set of experimental data. As we saw earlier, a general quantum channel Λ that transforms a state ρ to another state $\Lambda(\rho)$ can be represented by a series of Kraus operators $\{K_i\}$ such that:

$$\Lambda(\rho) = \sum_i K_i \rho K_i^\dagger. \tag{1.73}$$

In order to preserve the trace of the output state, the Kraus operators have to satisfy

$$\sum_i K_i^\dagger K_i = \mathbb{I}. \tag{1.74}$$

It is convenient to write Λ in a different form, which we achieve by first representing each Kraus operator as a linear combination of the Pauli operators[8] which we denote by $\{\sigma_1, \ldots, \sigma_{d^2}\}$:

$$K_i = \sum_{j=1}^{d^2} a_{ij}\sigma_j. \tag{1.75}$$

This allows us to rewrite Λ as [25]:

$$\Lambda(\rho) = \sum_{j=1}^{d^2}\sum_{k=1}^{d^2} \chi_{jk}\sigma_j\rho\sigma_k, \tag{1.76}$$

[8]We have switched to a slightly different notation than in Eq. (1.5). Here, the set of d-dimensional Pauli's $\{\sigma_1, \ldots, \sigma_{d^2}\}$ includes the identity matrix, and it forms a basis for d-dimensional Hermitian matrices.

where $\chi_{jk} = \sum_{i=1}^{d^2} a_{ij} a_{ik}^*$. This defines the *process matrix*, χ, which is what we want to reconstruct. The process matrix χ is a Hermitian, positive semi-definite matrix with size $d^2 \times d^2$ [5]. The constraint in Eq. (1.74) implies that

$$\sum_{j,k=1}^{d^2} \chi_{jk} \sigma_k \sigma_j = \mathbb{I}, \tag{1.77}$$

which defines d^2 constraints on χ. Thus, χ has a total of $d^4 - d^2 = d^2(d^2-1)$ free parameters, and since it is Hermitian and positive semidefinite we can parameterize it with the Cholesky decomposition $\chi = T^\dagger T$.

To perform process tomography, one subjects each of a tomographically complete set of states to the unknown transformation Λ, and then performs state tomography on the output, which requires data to be collected for at least d^2 measurement effects on each of d^2 states. Let ρ_a represent the a-th state, and E_b be the b-th measurement effect, and n_{ab} be the number of counts recorded when effect E_b is applied to the transformed state $\Lambda(\rho_a)$. The expected number of counts $\bar{n}_{ab}$ is given by $N \operatorname{Tr}\left(\sum_{j,k=1}^{d^2} \chi_{jk} \sigma_j \rho_a \sigma_k E_b \right)$. Again assuming that the measured counts n_{ab} follow a Gaussian distribution with mean and variance $\bar{n}_{ab}$, the probability of obtaining a string of measurement outcomes $\vec{n}$ is:

$$p(\vec{n}|\chi) = \frac{1}{N_{\text{norm}}} \exp\left\{ -\sum_{a,b} \frac{(n_{ab} - \bar{n}_{ab})^2}{2\bar{n}_{ab}} \right\}. \tag{1.78}$$

If we multiply both sides of equation (1.77) by σ_l and take the trace, we see that [33, 28] $\sum_{j,k=1}^{d^2} \operatorname{Tr}(\chi_{jk} \sigma_j \sigma_k \sigma_l) - \operatorname{Tr}(\sigma_l) = 0$, which defines one constraint for each of the d^2 σ_k's. These constraints can be enforced by use of a Lagrange multiplier, λ [28], such that the minimization problem we must solve for maximum likelihood quantum process tomography is:

$$\underset{\hat{\chi}}{\text{minimize}} : \sum_{a,b} \frac{\left[n_{ab} - N \operatorname{Tr}\left(\sum_{j,k=1}^{d^2} \hat{\chi}_{jk} \sigma_j \rho_a \sigma_k E_b \right) \right]^2}{N \operatorname{Tr}\left(\sum_{j,k=1}^{d^2} \hat{\chi}_{jk} \sigma_j \rho_a \sigma_k E_b \right)} + \lambda \sum_{l=1}^{d^2} \left(\sum_{j,k=1}^{d^2} \operatorname{Tr}(\hat{\chi}_{jk} \sigma_j \sigma_k \sigma_l) - \operatorname{Tr}(\sigma_l) \right)^2. \tag{1.79}$$

The process matrix $\hat{\chi}$ that solves the above problem is the maximum-likelihood estimate of the unknown quantum process.

1.5.4 Gate set tomography

The three tomography schemes presented above all require assumptions about some of the components in the experimental setup: state tomography requires a set of well-characterized measurements, and measurement tomography requires a set of well-characterized states. Process tomography has the most assumptions, as it requires accurate quantum descriptions of both states and measurements. If one has an experimental setup where neither the states nor measurements are well-characterized, a different tomography scheme is needed. One such scheme is known as *gate set tomography* [34, 35, 36, 37], which can simultaneously find maximum-likelihood estimates of a set of states, measurements, and quantum processes (or gates).

The setup is as follows: one has a single state preparation device which prepares unknown state ρ, and a single unknown measurement device with an unknown measurement effect E. In addition, one has a set of uncharacterized quantum processes $\mathcal{G}$ which can be individually applied to the system any number of times before it is measured. A given member, G_i, of the set $\mathcal{G}$ can be represented with the process matrix $\chi^{(G_i)}$, where $G_i(\rho) = \sum_{j,k=1}^{d^2} \chi_{jk}^{(G_i)} \sigma_j \rho \sigma_k$.

We identify a set $\mathcal{F} \subset \mathcal{G}$ whose elements allow us to transform our initial state ρ into each of a tomographically complete set of states. Then, to perform gate set tomography we must measure the quantities n_{ijk}, which are the number of outcomes obtained when the state ρ is subjected to transformation F_i, then G_j, and then F_k before being measured with the effect E. The expected number of outcomes is given by $\bar{n}_{ijk} = N \operatorname{Tr}\left(\sum_{mnstqr} \chi_{st}^{(F_k)} \chi_{qr}^{(G_j)} \chi_{mn}^{(F_i)} \sigma_s \sigma_q \sigma_m \rho \sigma_n \sigma_r \sigma_t E\right)$. If the quantity $\vec{n}$ encodes the data obtained after measuring n_{ijk} for all possible values of i, j, and k, then the probability of obtaining $\vec{n}$, given values for ρ, E, and $\mathcal{G}$ is:

$$p(\vec{n}|\rho, E, \chi^{(G_1)}, \ldots, \chi^{(G_g)}) = \frac{1}{N_{\text{norm}}} \exp\left\{ -\sum_{i,k=1}^{f} \sum_{j=1}^{g} \frac{(n_{ijk} - \bar{n}_{ijk})^2}{2\bar{n}_{ijk}} \right\}, \tag{1.80}$$

where we have assumed that $\mathcal{F}$ and $\mathcal{G}$ have f and g elements, respectively.

The maximum-likelihood estimates of the state, measurement, and transformations in the experiment is $\{\hat{\rho}, \hat{E}, \hat{\chi}^{(G_1)}, \ldots, \hat{\chi}^{(G_g)}\}$, and these are the solution to the following

minimization problem [36]:

$$\underset{\{\hat{\chi}^{(G_j)},\hat{\rho},\hat{E}\}}{\text{minimize}} : \sum_{i,k=1}^{f}\sum_{j=1}^{g} \frac{\left[n_{ijk} - N \operatorname{Tr}\left(\sum_{mnqrst} \hat{\chi}_{st}^{(F_k)} \hat{\chi}_{qr}^{(G_j)} \hat{\chi}_{mn}^{(F_i)} \sigma_s \sigma_q \sigma_m \hat{\rho} \sigma_n \sigma_r \sigma_t \hat{E}\right)\right]^2}{(\Delta n_{ijk})^2}, \tag{1.81}$$

$$\text{subject to} : \sum_{q,r=1}^{d^2} \operatorname{Tr}(\hat{\chi}_{qr}^{(G_i)} \sigma_q \sigma_r \sigma_l) - \operatorname{Tr}(\sigma_l) = 0 \quad \forall i \in \{1,\ldots,g\}, l \in \{1,\ldots,d^2\} \tag{1.82}$$

$$\operatorname{Tr}\hat{\rho} = 1, \tag{1.83}$$

$$\mathbb{I} - \hat{E} \geq 0. \tag{1.84}$$

We have assumed that $\hat{\rho}$, $\hat{E}$, and each $\hat{\chi}^{(G_i)}$ is parameterized with the Cholesky decomposition, which ensures these matrices are all positive semidefinite and Hermitian. The first set of constraints ensure that the constraint of Eq. (1.77) is satisfied for every $\hat{\chi}^{(G_i)}$, and the second constraints ensures that $\hat{\rho}$ is a valid density operator, and the third that $\hat{E}$ is a valid measurement effect.

Note also that the denominator of Eq. (1.81) is $(\Delta n_{ijk})^2$, instead of $\bar{n}_{ijk}$ as in Eq. (1.80). Δn_{ijk} is an estimate of the uncertainty in the measurement of n_{ijk}, and, if the maximum likelihood estimate fits the data well, $(\Delta n_{ijk})^2$ should be approximately equal to $\bar{n}_{ijk}$. The minimization problem for gate set tomography is more complicated than the corresponding ones in quantum state, measurement, or process tomography, and can be difficult to solve numerically. Using $(\Delta n_{ijk})^2$ instead of $\bar{n}_{ijk}$ simplifies the optimization without sacrificing too much accuracy [36].

Chapter 2

Hidden variable models of quantum theory

In Chapter 1 we saw that outcomes of projective quantum measurements are, in general, probabilistic in nature. (In fact, a projective measurement only has a deterministic outcome when the system being measured was prepared in an eigenstate of the measurement operator.) Quantum theory's explanation of this randomness is that upon encountering a measurement device, the system under measurement is instantly and probabilistically projected into one of the eigenstates of the operator describing the measurement. The words "instantly and probabilistically" make this explanation unsatisfactory. Classical intuition dictates that natural processes happen smoothly and continuously, and at first glance, the discontinuous jump from quantum state to measurement-operator-eigenstate is unsettling. Quantum theory accurately predicts the values of measurement outcome probabilities, but it offers no explanation of the underlying mechanism taking place.

Measurement-outcome statistics can only be gathered by repeating the same measurement on an ensemble of particles, and thus the interpretation of any apparent randomness should be restricted to such situations. One can easily imagine a situation in which each individual particle in an ensemble contains some pre-existing property which determines its measurement outcome, and that these properties are distributed among the members of the ensemble in such a way as to reproduce the predictions of quantum theory. These hypothetical pre-existing properties are sometimes called *hidden variables*. Models of this type are called *ontological*, because they assume that systems have real predetermined properties that specify their *ontic state*. In this thesis we will use the terms "hidden-variable model" and "ontological model" interchangeably.

We have just outlined two interpretations of what occurs when a quantum measurement takes place. Interpretation A simply accepts quantum theory's (lack of) explanation of the measurement procedure, and the corresponding discontinuous jump from state to measurement-eigenstate that goes along with it, and in interpretation B individual systems contain pre-existing properties described by hidden variables that we do not know how to measure. Neither interpretation is particularly appealing.

The possibility of an ontological model description of nature has been extensively investigated, and many theorems have been proven which show that any ontological model that makes the same predictions as quantum theory must sacrifice some other intuitive principle, such as noncontextuality (defined in Sections 2.2 and 2.3) or local causality (defined in Section 2.4). These are the theorems that will be presented and discussed in this chapter. We begin in Section 2.1 by introducing the ontological models framework. In Section 2.2 we review the notion of Bell-Kochen-Specker noncontextuality and proofs that quantum theory cannot be represented by ontological models that are Bell-Kochen-Specker noncontextual. In Section 2.3 we review a generalized definition of noncontexuality [38], and show that quantum theory cannot be represented by generalised noncontextual ontological models either. We end in Section 2.4 with a review of Bell's theorem [39], which proves that locally causal hidden-variable models cannot reproduce the predictions of quantum theory.

2.1 The ontological models framework

We will begin by introducing the ontological models framework.

An ontological model of an operational theory models a preparation device P as one that prepares a system in an ontic state λ according to a probability distribution $\mu(\lambda|P)$ (Fig. 2.1(a)). Because μ is a probability distribution, we have $\mu(\lambda|P) \geq 1$ and $\int d\lambda \mu(\lambda|P) = 1$.

A K-outcome measurement device, M, with outcomes $m(k)$ for $k \in \{1, \ldots, K\}$, is represented in the ontological model by a set of response functions $\xi_{m(k)}(\lambda|M)$. A measurement samples the system's ontic state λ and returns an outcome $m(k)$ with probability $\xi_{m(k)}(\lambda|M)$. Since a measurement has to return exactly one outcome, we have $\sum_{k=1}^{K} \xi_{m(k)}(\lambda|M) = 1$, and since the response functions represent probabilities, $\xi_{m(k)}(\lambda|M) \geq 0$ for all $m(k)$, λ, and M. Figure 2.1(b,c) depicts possible response functions for two different measurements.

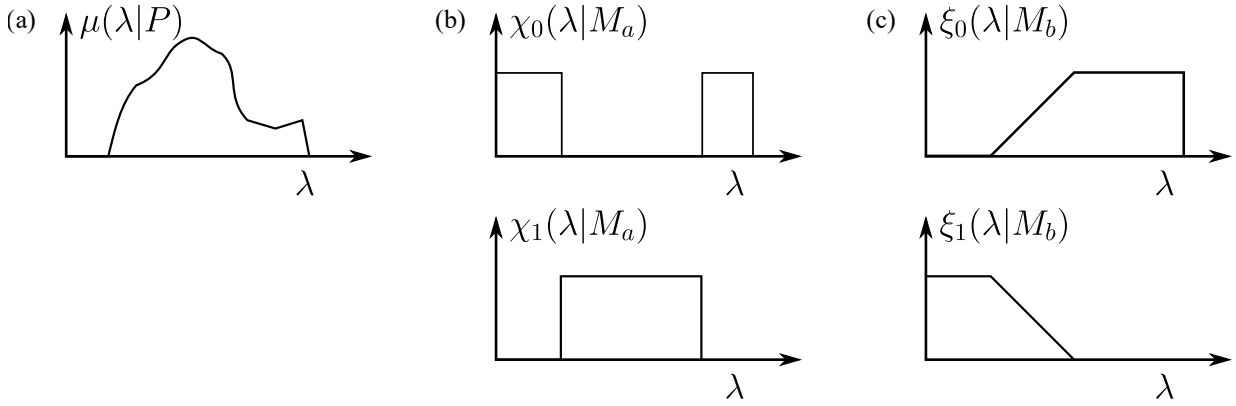

Figure 2.1: Representation of preparation and measurement procedures in an ontological model. (a) The preparation procedure P prepares a system in the state λ according to probability distribution $\mu(\lambda|P)$. (b) Response functions for measurement, M_a, with two outcomes $\{0, 1\}$. The response functions of M_a are deterministic. (c) A measurement, M_b, with two outcomes $\{0, 1\}$ and nondeterministic response functions.

The probability of obtaining outcome $m(k)$ when measurement procedure M is performed on a system that was prepared with procedure P is:

$$p\left(m(k)|P, M\right) = \int d\lambda \mu\left(\lambda|P\right) \xi_{m(k)}(\lambda|M). \tag{2.1}$$

2.1.1 Outcome-deterministic ontological models

Later we will be concerned with a special class of ontological models, namely ones for which all measurements respond deterministically to the ontic state λ.

A deterministic measurement procedure is represented by a set of deterministic response functions $\chi_k(\lambda|M)$ (Fig. 2.1(b)). These response functions can only take the values 0 or 1, and for a given λ, the set of response functions describing the measurement equal 1 for exactly one value of k, and 0 for all others. The probability of obtaining outcome k when performing measurement M on preparation P is:

$$p(m(k)|P, M) = \int d\lambda p(\lambda|P) p(m(k)|\lambda, M), \tag{2.2}$$

$$= \int d\lambda \mu(\lambda|P) \chi_{m(k)}(\lambda|M). \tag{2.3}$$

An ontological model for which all measurements are deterministic is an *outcome-deterministic ontological model* (ODOM).

An alternative way to understand the outcome-deterministic ontological model is to think of every ontic state λ as having preassigned outcomes for each possible measurement which might be performed on the system. Since measurements respond deterministically to the ontic state, choosing a system to be in a specific state λ is equivalent to assigning outcomes for any future measurements.

2.1.2 General features of ontological models

There are a couple of features of ontological models which will be useful later on. We will list them here.

- *Distinguishable preparation procedures cannot have overlapping distributions in the ontological model.*

  Consider two preparation procedures P and P' which can be perfectly distinguished by a single-shot measurement. That implies that P and P' cannot have overlapping distributions in the ontological model, that is:

  $$\mu(\lambda|P)\,\mu(\lambda|P') = 0. \tag{2.4}$$

  This must be true based on the ontological model description of measurements. A measurement samples the ontic state λ, then returns an outcome based on the ontic state. If P and P' had overlapping distributions in the ontological model, then there would exist values of λ consistent with both P and P', and no measurement would be able to perfectly distinguish between the two preparation procedures for these values of λ.

- *Convex combinations of experimental procedures are represented with convex mixtures in the ontological model.*

  Consider a convex mixture of the preparation procedures P and P'. P_{mix} is the procedure that, with probability w, implements procedure P, and with probability $1-w$ implements procedure P'. The probability of finding the system in a specific ontic state λ is thus the probability that P was implemented, multiplied by the probability of finding the system in state λ given that P was implemented, plus the

probability that P' was implemented, multiplied by the probability of finding the system in state λ given that P' was implemented. Hence,

$$\mu\left(\lambda|P_{\text{mix}}\right) = w\mu\left(\lambda|P\right) + (1-w)\mu\left(\lambda|P'\right) \tag{2.5}$$

Convex mixtures of measurements are represented in a similar way. For example, assume that one has two K-outcome measurement procedures M and M', with sets of outcomes $\{m(k) = k\}$ and $\{m'(k) = k\}$, respectively. Let M_{mix} be the measurement procedure that, with probability w, implements measurement M, and outputs outcome $m_{\text{mix}}(m(k)) = m_{\text{mix}}(k)$, and, with probability $(1-w)$, implements procedure M' and returns outcome $m_{\text{mix}}(m'(k)) = m_{\text{mix}}(k)$. In the ontological model, the response functions of M_{mix} are:

$$\xi_{m(k)}(\lambda|M_{\text{mix}}) = w\xi_{m(k)}(\lambda|M) + (1-w)\xi_{m'(k)}(\lambda|M'). \tag{2.6}$$

2.2 The Bell-Kochen-Specker theorem

The Bell-Kochen-Specker (BKS) theorem (independently discovered by Kochen and Specker [40] and Bell [41]) concerns outcome-deterministic ontological models. The BKS theorem states that no noncontextual, outcome-deterministic ontological model can reproduce the predictions of quantum theory. We will begin this section by motivating the attempt to represent quantum theory with an ODOM, then define the BKS notion of noncontextuality, and finally give and discuss a proof of the BKS theorem.

2.2.1 PVMs and ODOMs

Consider a rank-1 projective quantum measurement (PVM) A defined by a set of K rank-1 projective effects $\{E_1^{(A)}, \ldots, E_K^{(A)}\}$ with corresponding outcomes $\{1, \ldots, K\}$. We define each effect as $E_k^{(A)} = |\psi_k\rangle\langle\psi_k|$, satisfying $\langle\psi_j \mid \psi_k\rangle = \delta_{j,k}$. The PVM responds deterministically to the set of basis states $\{|\psi_1\rangle, \ldots, |\psi_K\rangle\}$, i.e. $p\left(k|\psi_j, A\right) = \text{Tr}(|\psi_j\rangle\langle\psi_j| E_k^{(A)}) = \delta_{j,k}$.

For an ensemble of systems all identically prepared in the state ρ, the probability of

obtaining outcome k when performing measurement A is:

$$p\left(k|\rho,A\right)=\sum_{j=1}^{K}p\left(\psi_j|\rho\right)p\left(k|\psi_j,A\right), \tag{2.7}$$

$$=\sum_{j=1}^{K}\mathrm{Tr}\left(\rho|\psi_j\rangle\langle\psi_j|\right)\delta_{j,k}. \tag{2.8}$$

Here, $p\left(\psi_j|\rho\right)=\mathrm{Tr}\left(\rho|\psi_j\rangle\langle\psi_j|\right)$ represents the probability that the measurement A will project ρ into eigenstate $|\psi_j\rangle$, and $p\left(k|\psi_j,A\right)=\delta_{j,k}$ is the probability that outcome k will be returned given that ρ was projected into state $|\psi\rangle$.

There is a clear analogy between a projective measurement in quantum theory and a deterministic measurement in the ontological models framework, which can most easily be seen by comparing equations (2.7) and (2.2). In both equations, the probability of finding the system in a specific (quantum or ontic) state is calculated, then multiplied by the probability of returning outcome k for that state. These probabilities are then integrated over all possible states that the system could be found to be in. In the quantum description, projective measurements respond deterministically to eigenstates of their measurement operators, and in the ODOM measurements respond deterministically to the ontic state λ. For now, *we will assume* that the ontological model representation of a projective quantum measurement is deterministic. We will revisit this assumption (and provide additional justification for it) in Section 2.3.3.

Compatible measurements

Consider two commuting PVMs, A and B. Quantum theory tells us that A and B share a common set of K measurement effects, $\{E_k\}$, with $E_k=|\psi_k\rangle\langle\psi_k|$ for some basis $\{|\psi_k\rangle\}$. Since we don't have to label the outcomes of our two measurements in the same way, we will say that A has a set of outcomes $\{a(k)\}$ and B has a set of outcomes $\{b(k)\}$, where outcomes $a(k)$ and $b(k)$ correspond to effect E_k. We call the pair $(a(k),b(k))$ a set of *simultaneous outcomes*, because in a measurement of A and B these outcomes will always be returned in pairs. Hence, we must have:

$$p\left(a(j)|\psi_k,A\right)=p\left(b(j)|\psi_k,B\right)=\delta_{j,k}. \tag{2.9}$$

In other words, the results of a simultaneous measurement of commuting observables must be a corresponding set of simultaneous outcomes [42]. This further restricts the assignable

values in the ODOM. To ensure agreement with quantum theory, the response functions in an ODOM are constrained in an identical way. Specifically, for two compatible measurement procedures M_A and M_B (representing compatible PVMs with operators A and B), the response functions in the ODOM must satisfy:

$$\chi_{a(k)}(\lambda|M_A) = \chi_{b(k)}(\lambda|M_B) \in \{0, 1\}. \tag{2.10}$$

In other words, for each specific value of the ontic state λ, the preassigned outcomes for a set of compatible observables must be a corresponding set of simultaneous outcomes.

2.2.2 Bell-Kochen-Specker noncontextuality

The discussion of compatible measurements above contained an implicit assumption which is worth making explicit. This assumption can most easily be seen by considering three projective measurements with observables A, B, and C. Suppose that observables A and B commute, A and C also commute, but B and C do not. This means that A can be jointly measured with either B, or with C, but not with both. These two joint measurements (i.e. A measured with B and A measured with C) determines two different *measurement contexts* for A.

The implicit assumption in our ODOM is also called the *assumption of BKS noncontextuality*, and it is simply this: in an outcome-deterministic ontological model, the response functions describing a measurement procedure do not depend on the context in which it is measured.

The assumption of BKS noncontextuality seems very reasonable. After all, in quantum theory we use the same set of measurement effects to represent observable A regardless of if we measure A jointly with B or C. There is no *a priori* reason to expect the ontological model description of A to depend on its measurement context. However, as we are about to see, the BKS theorem proves that no noncontextual outcome-deterministic ontological model can reproduce all the predictions of quantum mechanics.

2.2.3 Peres-Mermin square proof of the BKS theorem

The original proofs of the BKS theorem by Kochen and Specker [40] and Bell [41] are similar in that both proofs work by first finding a specific set of projective measurement operators, then showing that one always finds a contradiction if they try to describe the measurement outcomes with a noncontextual outcome-deterministic ontological model.

$\sigma_x \otimes \mathbb{I}$	$\mathbb{I} \otimes \sigma_x$	$\sigma_x \otimes \sigma_x$	$\mathbb{I} \otimes \mathbb{I}$
$\mathbb{I} \otimes \sigma_y$	$\sigma_y \otimes \mathbb{I}$	$\sigma_y \otimes \sigma_y$	$\mathbb{I} \otimes \mathbb{I}$
$\sigma_x \otimes \sigma_y$	$\sigma_y \otimes \sigma_x$	$\sigma_x \otimes \sigma_z$	$\mathbb{I} \otimes \mathbb{I}$
$\mathbb{I} \otimes \mathbb{I}$	$\mathbb{I} \otimes \mathbb{I}$	$-\mathbb{I} \otimes \mathbb{I}$	

Table 2.1: Set of projective measurements for the Peres–Mermin magic square proof of the BKS theorem. The upper-left 3×3 portion of the table represents the nine projective measurements needed for Mermin's proof. The rightmost column represents the products of the observables in each row, and the bottom row is the products of the observables in each column.

The original proofs are somewhat complicated and are outside the scope of this thesis. For the sake of brevity and clarity we will present a simpler alternative proof due to Mermin and Peres [43, 44, 45, 42].

Consider the set of projective measurement operators in Table 2.1. We will be considering measurements of the nine observables in the upper-left 3×3 corner of the table. The measurements are performed on a four-level quantum system. The eigenvalues (and thus possible outcomes) of each observable are ± 1. One can check that the each of the three observables in each row (column) commute with the others in the row (column). Thus, each row and column is a set of three compatible measurements which can be measured jointly. Each observable is a member of two possible measurement contexts, since it may be jointly measured with the other two observables in its row, or it may be jointly measured with the other two observables in its column.

A joint measurement of each of the observables in a row or column also implies a measurement of a fourth observable equal to the product of the three being measured. The observables in the fourth row (column) are the products of the three observables in each column (row). The product of the three observables in each row and column is $\mathbb{I} \otimes \mathbb{I}$, except for the rightmost column which has product $-\mathbb{I} \otimes \mathbb{I}$. This implies that in a joint measurement of each row and two of the columns the product of the outcomes must be $+1$, and in a joint measurement of the rightmost column the product of the outcomes must be -1.

Our task now is to try to find a noncontextual ODOM that can reproduce this set of quantum correlations. We need to specify a set of response functions for each of the nine observables in the upper-left 3×3 corner of Table 2.1, or equivalently, for each value that λ can take, we need to specify an outcome for each observable. Each of these measurements can return either $+1$ or -1 as an outcome, and thus these are the outcomes

we can preassign. However, the products of the three outcomes in each row and the first two columns are constrained to be 1, and the product of the three outcomes in the third column must be -1. The row constraints require an even number of -1 outcomes in each row (and thus in the table as a whole), while the column constraints require an odd total number of -1 outcomes. These two constraints contradict each other, and thus we conclude that no noncontextual outcome-deterministic ontological model can reproduce all the predictions of theory.

In addition to the works due to Bell [41], Kochen and Specker [40], and Peres and Mermin [43, 44, 45, 42], there are many other proofs of the BKS theorem; two notable examples are due to Cabello [46] and Klyachko, Can, Binicioğlu and Shumovsky [47]. For brevity, we will omit a review of all approaches to proving the BKS theorem. However, we note that these works all prove the same point: projective quantum measurements cannot be represented by an ontological model that is both noncontextual and outcome-deterministic.

2.2.4 Can the BKS theorem be confirmed experimentally?

The BKS theorem proves that no noncontextual outcome-deterministic ontological model can reproduce all predictions of quantum theory. It is a remarkable theorem, and it raises the obvious question of whether or not it can be confirmed experimentally. Or, more precisely, the question of whether or not one can perform an experiment proving that *nature* (and not just quantum theory) cannot be represented by an outcome-deterministic ontological model that is noncontextual in the Bell-Kochen-Specker sense. The answer, strictly speaking, is that the BKS theorem (as presented) *cannot* be confirmed experimentally, and we explain in this section why this is the case.

The BKS theorem is a mathematical statement about the interpretations of quantum mechanics, and doesn't say anything about the physical theory describing nature itself. It is possible that quantum theory isn't 100% correct, and that one day it will be surpassed by a new physical theory which describes nature more accurately. Still, one can imagine that whatever physical theory actually describes nature contains a set of measurements that give the same measurement-outcome statistics as the observables in Table 2.1. If this were the case, one could further imagine performing an experiment which confirms the quantum predictions for the outcomes of these measurements. Surely such an experiment would be sufficient evidence that the physical theory describing nature could not be interpreted as a noncontextual outcome-deterministic ontological model. However, even in this hypothetical scenario we are immediately confronted with two experimental difficulties, which can be understood by carefully considering what such an experiment would entail.

The problem of inexact operational equivalence

To elucidate the first difficulty, we will use the Peres-Mermin square as an example.

In order to apply the conclusions of the Peres-Mermin proof to a set of experimental data, an experimentalist would have to build or find a set of nine measurement devices and confirm that their outcome statistics are consistent with the predictions made by quantum theory. This would require one to gather measurement-outcome statistics on each of the six joint measurements (three row measurements and three column measurements) in Table 2.1. This requires performing each measurement in the table twice, once in each measurement context it appears in (i.e. in its row and column contexts), requiring a rearrangement of the experimental set-up in some way, whether that be changing parameters of some measurement device, or swapping different devices in and out of the experiment. In our noncontextual ODOM, we restricted the outcome assigned to each observable to be independent of which context it was measured in. However, if we have to change the experimental set-up in order to measure an observable in two different contexts, how can we be sure that we are measuring the *same* observable in each context? If we aren't sure that we are measuring the same observable in each context, we have no reason to assume that it should respond in the same way to the ontic state of the system, thus making it unreasonable to apply our assumption of BKS noncontextuality[1].

An obvious solution is simply to check that a measurement device has the same performance in both contexts. For example, consider two sets of compatible measurements $\{A, B\}$ and $\{A', C\}$, and our goal is to verify that A and A' truly are the same measurement. We can imagine sending many copies of every possible state into measurement devices A and A' and collecting statistics on the measurement outcomes. If the outcome statistics for A equal those for A' for every input state, then we say that A and A' are *operationally equivalent* (an exact definition of operational equivalence follows in Section 2.3.1), and we are free to apply the assumption of noncontextuality and represent A and A' in the ontological model with the same deterministic measurement. However, systematic errors (which are *always* present in experiments) will make it impossible to implement a joint measurement of A and B and a joint measurement of A' and C so precisely that A and A' are exactly operationally equivalent. While A and A' might be operationally "close" to one another, they will never be exactly equivalent, and thus we will have no warrant to apply the assumption of BKS noncontextuality to these measurements. This is the first

[1]This was in fact one of Bell's criticisms of his own proof of the BKS theorem [41], and his solution was to consider multipartite systems in which compatible measurements could be separated in a space-like way, thus replacing the assumption of noncontextuality with a perhaps more defensible assumption of local causality. This idea will be expanded upon in Section 2.4.

experimental difficulty we encounter, and we call this the *problem of inexact operational equivalence*.

The problem of noisy measurements

The second experimental difficulty is also related to the unavoidable presence of noise in realistic experiments. Our formulation of the ODOM was based on the quantum description of projective measurements. Projective measurements respond deterministically to the eigenstates of the operator representing the measurement — this is why we restricted ourselves to only consider ontological models which respond deterministically to a system's ontic state. However, in quantum theory, any measurement with a finite amount of noise *does not* have a set of states to which it responds deterministically, and thus it does not make sense to model noisy quantum measurements with an ODOM. All realistic experiments will contain some amount of noise, and thus ontological models that include *indeterministic* measurements should be considered as valid possible explanations of these experiments as well. The BKS theorem only considers ODOMs, and therefore data obtained with noisy measurements cannot be used to support it. This is the second difficulty one encounters in experimental tests of noncontextuality, and we call it the *problem of noisy measurements*.

2.3 Generalised noncontextuality for preparations and measurements

To address the problem of noisy measurements, Spekkens [38] generalised the notion of BKS noncontextuality so it can be applied to all ontological models, as opposed to only outcome-deterministic ones. In this section we introduce generalised noncontextuality, and show that quantum theory does not admit of a generalised noncontextual ontological model.

2.3.1 Generalised preparation and measurement noncontextuality

Generalised noncontextuality applies to pairs of experimental procedures which cannot be distinguished from each other [38]. In words, the *assumption of generalised noncontextuality* is this: *two experimental procedures that are operationally equivalent* (a notion

that will be precisely defined below) *should be represented identically in the underlying model*. We say "experimental procedures" instead of "measurement procedures" because generalised noncontextuality can be applied to state preparation procedures in addition to measurement procedures[2].

If one considers just measurements, it is clear how generalised noncontextuality is a generalisation of BKS noncontextuality. As discussed in Section 2.2.4, the assumption of BKS noncontextuality states that an observable's response function in an ODOM cannot depend on which context it is performed in. In other words, since it is the same observable being measured in each context, it should be represented the same way in the ODOM regardless of which context it is in. Experimentally, the only way to check if a measurement device measures the same observable in two contexts is to measure its outcome statistics for all preparations in both contexts: if the statistics are different we conclude that the device behaves differently in each context, but if the statistics are the same we can conclude the observables being measured in each context are identical. Thus, the assumption of BKS noncontextuality can only be applied to measurement devices whose operation doesn't depend on context. Spekkens' notion of noncontextuality simply generalises this definition to include noisy measurement devices and preparation devices.

Before proceeding with a more formal definition of Spekkens noncontextuality, we first formally review the notion of *operational equivalence*.

Operational equivalence

Consider two preparation procedures, P and P'. We say that P and P' are *operationally equivalent* (denoted by $P \approx P'$) if there is no measurement that we can perform which can distinguish a system prepared with P from one prepared with P'. If $\mathcal{M}$ is the set of all possible measurements that can be performed, then mathematically, we say that $P \approx P'$ if

$$p\left(m(k)|P, M\right) = p\left(m(k)|P', M\right), \quad \forall M \in \mathcal{M}, \quad \forall k \in \{1, \ldots, K\} \tag{2.11}$$

Similarly, two measurement procedures are operationally equivalent if there exists no preparation procedure for which the two measurements return differing outcome statistics. Mathematically, if M and M' are two K-outcome measurement procedures, and if $\mathcal{P}$ is the set of all accessible preparation procedures, then we say that M and M' are operationally equivalent (i.e. we say that $M \approx M'$) if

$$p\left(m(k)|P, M\right) = p\left(m'(k)|P, M'\right), \quad \forall P \in \mathcal{P}, \quad \forall k \in \{1, \ldots, K\}. \tag{2.12}$$

[2]In fact, when Spekkens introduced the assumption of generalised noncontextuality he also defined it for state transformations [38], although we will not discuss this in this thesis.

Generalised noncontextual ontological models

The assumption of generalised noncontextuality can be applied to both preparation and measurement procedures.

In an ontological model, the *assumption of generalised preparation noncontextuality* states that equivalence of preparations at the operational level implies equivalence of the preparations at the ontological level:

$$P \approx P' \implies \mu(\lambda|P) = \mu(\lambda|P'). \tag{2.13}$$

In an ontological model, the *assumption of generalised measurement noncontextuality*, states that equivalence of measurement procedures at the operational level implies equivalence of the measurements at the ontological level:

$$M \approx M' \implies \xi(k(M)|\lambda, M) = \xi(k(M')|\lambda, M'), \quad \forall k \in \{1, \ldots, n\}. \tag{2.14}$$

Throughout the rest of this thesis, we will use the terms "preparation noncontextuality" and "measurement noncontextuality" as shorthand for "the assumption of generalised preparation noncontextuality" and "the assumption of generalised measurement noncontextuality". When referring to both preparation and measurement noncontexuality we will use the term "generalised noncontextuality", or sometimes just "noncontextuality".

It is immediately obvious that the notion of generalised noncontextuality has some advantages over the notion of BKS noncontextuality. First, we can apply the assumption of generalised noncontextuality to outcome-indeterministic ontological models, which solves the problem of noisy measurements discussed in Section 2.2.4. Second, it is more general, as it can be applied to preparations as well as measurements. Third, as we will see, one can use qubit states and measurements to prove that quantum theory does not admit of a generalised noncontextual model, whereas Hilbert spaces of dimension at least three are required to show that quantum theory is inconsistent with a BKS-noncontextual ontological model.

2.3.2 Qubits cannot be represented by a preparation noncontextual ontological model

After introducing the notion of generalised noncontextuality, Spekkens proved that a qubit cannot be represented by a preparation noncontexual ontological model [38]. We reproduce his argument in this section.

Consider the six preparation procedures $P_{1,0}$, $P_{1,1}$, $P_{2,0}$, $P_{2,1}$, $P_{3,0}$ and, $P_{3,1}$, which are used to prepare the following quantum states:

$$|\psi_{1,0}\rangle = |0\rangle \tag{2.15}$$
$$|\psi_{1,1}\rangle = |1\rangle \tag{2.16}$$
$$|\psi_{2,0}\rangle = \frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle \tag{2.17}$$
$$|\psi_{2,1}\rangle = \frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle \tag{2.18}$$
$$|\psi_{3,0}\rangle = \frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle \tag{2.19}$$
$$|\psi_{3,1}\rangle = \frac{\sqrt{3}}{2}|0\rangle + \frac{1}{2}|1\rangle\,. \tag{2.20}$$

Each pair of preparations $\{P_{i,0}, P_{i,1}\}$ corresponds to a pair of orthogonal quantum states, i.e.

$$\langle\psi_{1,0}|\psi_{1,1}\rangle = 0 \tag{2.21}$$
$$\langle\psi_{2,0}|\psi_{2,1}\rangle = 0 \tag{2.22}$$
$$\langle\psi_{3,0}|\psi_{3,1}\rangle = 0 \tag{2.23}$$

These six pure states lie around the circumference of the XZ plane of the Bloch sphere, depicted in Fig. 2.2. Now consider the preparation P_i which is the procedure that implements one of $P_{i,0}$ or $P_{i,1}$, the choice of which is made uniformly at random. Further consider P_4 which implements one of $P_{1,0}$, $P_{2,0}$, or $P_{3,0}$ uniformly at random, and P_5 which implements one of $P_{1,1}$, $P_{2,2}$, or $P_{3,3}$ uniformly at random. Each of these five convex mixtures of preparation procedures prepares the maximally mixed state, which is represented with same density operator:

$$\frac{\mathbb{I}}{2} = \frac{1}{2}|\psi_{1,0}\rangle\langle\psi_{1,0}| + \frac{1}{2}|\psi_{1,1}\rangle\langle\psi_{1,1}| \tag{2.24}$$
$$= \frac{1}{2}|\psi_{2,0}\rangle\langle\psi_{2,0}| + \frac{1}{2}|\psi_{2,1}\rangle\langle\psi_{2,1}| \tag{2.25}$$
$$= \frac{1}{2}|\psi_{3,0}\rangle\langle\psi_{3,0}| + \frac{1}{2}|\psi_{3,1}\rangle\langle\psi_{3,1}| \tag{2.26}$$
$$= \frac{1}{3}|\psi_{1,0}\rangle\langle\psi_{1,0}| + \frac{1}{3}|\psi_{2,0}\rangle\langle\psi_{2,0}| + \frac{1}{3}|\psi_{3,0}\rangle\langle\psi_{3,0}| \tag{2.27}$$
$$= \frac{1}{3}|\psi_{1,1}\rangle\langle\psi_{1,1}| + \frac{1}{3}|\psi_{2,1}\rangle\langle\psi_{2,1}| + \frac{1}{3}|\psi_{3,1}\rangle\langle\psi_{3,1}|\,, \tag{2.28}$$

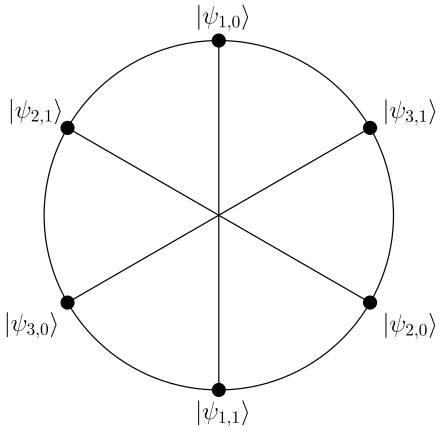

Figure 2.2: Six states required for proof that a qubit cannot be represented by a preparation noncontextual model. These states lie on the X-Z plane of the Bloch sphere. They are defined by Eqs. (2.15)–(2.20), and they satisfy the operational equivalences of Eqs. (2.24)–(2.28)

.

and thus quantum theory predicts that these five convex mixtures will give rise to the same measurement-outcome statistics.

The preparation procedures $P_{i,0}$ and $P_{i,1}$ prepare orthogonal quantum states, and thus their representations in the ontological model must be non-overlapping (Eq. (2.4)). This implies:

$$\mu\left(\lambda|P_{1,0}\right)\mu\left(\lambda|P_{1,1}\right)=0 \tag{2.29}$$
$$\mu\left(\lambda|P_{2,0}\right)\mu\left(\lambda|P_{2,1}\right)=0 \tag{2.30}$$
$$\mu\left(\lambda|P_{3,0}\right)\mu\left(\lambda|P_{3,1}\right)=0. \tag{2.31}$$

The preparation procedures $P_1,\ldots,P_5$ are convex mixtures, and thus their representations in the ontological model can be represented as convex mixtures as well (Eq. (2.5)). Furthermore, since quantum theory predicts that $P_1,\ldots,P_5$ are operationally equivalent (Eqs. (2.24)-(2.28)), preparation noncontextuality dictates that $\mu\left(\lambda|P_1\right)=\cdots=\mu\left(\lambda|P_5\right)$.

Thus, if we let $\nu(\lambda) = \mu(\lambda|P_1) = \cdots = \mu(\lambda|P_5)$, we have:

$$\nu(\lambda) = \frac{1}{2}\mu(\lambda|P_{1,0}) + \frac{1}{2}\mu(\lambda|P_{1,1}) \tag{2.32}$$

$$= \frac{1}{2}\mu(\lambda|P_{2,0}) + \frac{1}{2}\mu(\lambda|P_{2,1}) \tag{2.33}$$

$$= \frac{1}{2}\mu(\lambda|P_{3,0}) + \frac{1}{2}\mu(\lambda|P_{3,1}) \tag{2.34}$$

$$= \frac{1}{3}\mu(\lambda|P_{1,0}) + \frac{1}{3}\mu(\lambda|P_{2,0}) + \frac{1}{3}\mu(\lambda|P_{3,0}) \tag{2.35}$$

$$= \frac{1}{3}\mu(\lambda|P_{1,1}) + \frac{1}{3}\mu(\lambda|P_{2,1}) + \frac{1}{3}\mu(\lambda|P_{3,1}) \tag{2.36}$$

We can now show that, aside from the trivial "all-zero" solution $0 = \mu(\lambda|P_{1,0}) = \cdots = \mu(\lambda|P_{3,1})$, it is impossible to find distributions $\mu(\lambda|P_{1,0}) \ldots \mu(\lambda|P_{3,1})$ which solve Eqs. (2.29)-(2.36) simultaneously. First, note that Eq. (2.29) requires that, for each value of λ one of $\mu(\lambda|P_{1,0})$ and $\mu(\lambda|P_{1,1})$ must be zero, and similarly, Eqs. (2.30) and (2.31) require that one of $\{\mu(\lambda|P_{2,0}), \mu(\lambda|P_{2,1})\}$ and one of $\{\mu(\lambda|P_{3,0}), \mu(\lambda|P_{3,1})\}$, must be zero. There are eight possible arrangements of zeros which satisfy Eqs. (2.29)-(2.31).

If we choose $\mu(\lambda|P_{1,0}) = \mu(\lambda|P_{2,0}) = \mu(\lambda|P_{3,0}) = 0$, then Eqs. (2.35) and (2.36) imply that $\mu(\lambda|P_{1,1})$, $\mu(\lambda|P_{2,1})$, and $\mu(\lambda|P_{3,1})$ are all zero as well. If instead we choose $\mu(\lambda|P_{1,1}) = \mu(\lambda|P_{2,0}) = \mu(\lambda|P_{3,0}) = 0$, then Eqs. (2.32) and (2.35) imply that $\frac{1}{2}\mu(\lambda|P_{1,0}) = \frac{1}{3}\mu(\lambda|P_{1,0})$, which in turn implies that $\mu(\lambda|P_{1,0}) = 0$. But now we again have $\mu(\lambda|P_{1,0}) = \mu(\lambda|P_{2,0}) = \mu(\lambda|P_{3,0}) = 0$, and so we conclude that all distributions are zero. Considering each of the other six arrangements of zeros that satisfy Eqs. (2.29)-(2.31) will lead us to the same conclusion. Hence, for every choice of λ, we are lead to the trivial solution $0 = \mu(\lambda|P_{1,0}) = \cdots = \mu(\lambda|P_{3,1})$. Since the trivial solution is not a normalized probability distribution, we conclude that no preparation noncontextual ontological model can describe this set of qubit states.

2.3.3 Measurements on qubits cannot be represented by a generalised noncontextual ontological model

Up until this point in the thesis, we have assumed that projective quantum measurements should be represented with deterministic response functions in the ontological model. Spekkens proves that [38], under this condition, no measurement noncontextual ontological can describe the measurements that can be performed on a qubit.

Spekkens has also shown that that [38], under the assumption of preparation noncontextuality, projective quantum measurements *must* be deterministic in the ontological model. In this section, we will first argue that preparation noncontextuality implies that an ontological model of a PVM must have deterministic response functions, and then we will proceed with the proof that qubit measurements cannot be represented by a generalised noncontextual ontological model.

Preparation noncontextuality implies outcome determinism for projective-valued measures

Here we sketch an argument supporting the claim that preparation noncontextuality implies that an ontological model of a quantum PVM must be outcome deterministic. The interested reader can find full details of this proof in Appendix A of Ref. [38].

Now we begin the sketch of the argument. In quantum mechanics, a projective-valued measure M is defined by a set of rank-1 measurement effects $\{E_k\}$. The PVM effects can be used to define a basis $\{|\psi_j\rangle\}$ such that $\mathrm{Tr}(|\psi_j\rangle\langle\psi_j| E_k) = \delta_{j,k}$. Let state $|\psi_k\rangle$ be that prepared by procedure P_k. In the ontological model, the distributions $\mu(\lambda|P_k)$ must be nonoverlapping (Eq. (2.4)), and therefore, their regions of support Ω_k (defined as $\Omega_k = \{\lambda : \mu(\lambda|P_k) > 0\}$) must be disjoint. In order to agree with quantum theory, the response functions must also satisfy $\int d\lambda \mu(\lambda|P_j)\xi_k(\lambda|M) = \delta_{j,k}$, which implies that $\xi_k(\lambda|M) = 1$ for all $\lambda \in \Omega_k$ and $\xi_k(\lambda|M) = 0$ for all $\lambda \in \Omega_{j\neq k}$. Thus, for every $\lambda \in \bigcup_k \Omega_k$, exactly one response function is equal to 1, and all the others are 0, and hence M's response functions are deterministic for all λ in support of the full set of quantum states $\{|\psi_k\rangle\}$.

The next half of the argument demonstrates that the full ontic state space Ω is in fact equal to the union of the supports of the set of quantum basis states, $\lambda \in \Omega = \bigcup_k \Omega_k$, which implies that the response functions are deterministic everywhere. If $P_{\mathbb{I}/d}$ is the procedure that creates the maximally-mixed state $\mathbb{I}/d$ then we can define Ω as the set of ontic states for which $\mu(\lambda|P_{\mathbb{I}/d}) > 0$, since every possible quantum state is part of some convex decomposition of the maximally-mixed state. Now, note that $\sum_k |\psi_k\rangle\langle\psi_k| = \mathbb{I}$. Applying the assumption of preparation noncontextuality implies that $\sum_k \frac{1}{d}\mu(\lambda|P_k) = \mu(\lambda|P_{\mathbb{I}/d})$, which in turn implies that $\Omega = \bigcup_k \Omega_k$.

Proof that qubit measurements cannot be represented by a generalised noncontextual ontological model

Now we are ready to show that a general quantum measurement cannot be described by a noncontextual ontological model, which was first shown in Ref. [38].

We will consider three measurements, M_1, M_2, and M_3, represented by projective-valued measures $\{E_{1,0}, E_{1,1}\}$, $\{E_{2,0}, E_{2,1}\}$, and $\{E_{3,0}, E_{3,1}\}$. For each $E_{i,j}$, the first index i represents the measurement that the projector is part of, and the second index j is what we label the outcome corresponding to that projector. Each measurement is defined to respond deterministically to one of the pairs of orthogonal states in Eqs. (2.15)-(2.20). Specifically, we have

$$\mathrm{Tr}\left(|\psi_{i,j}\rangle\langle\psi_{i,j}| E_{i,j}\right) = 1 \quad \text{for all } i \in \{1,2,3\} \text{ and } j \in \{0,1\}. \tag{2.37}$$

The assumption of preparation noncontextuality implies that, in the ontological model, M_i is represented by the pair of outcome-deterministic response functions $\{\chi_0(\lambda|M_i), \chi_1(\lambda|M_i)\}$, for $i \in \{1,2,3\}$. By definition of an outcome-deterministic measurement in an ontological model, these response functions must satisfy:

$$\chi_0(\lambda|M_1) + \chi_1(\lambda|M_1) = 1 \tag{2.38}$$
$$\chi_0(\lambda|M_2) + \chi_1(\lambda|M_2) = 1 \tag{2.39}$$
$$\chi_0(\lambda|M_3) + \chi_1(\lambda|M_3) = 1, \tag{2.40}$$

as well as

$$\chi_0(\lambda|M_1)\chi_1(\lambda|M_1) = 0 \tag{2.41}$$
$$\chi_0(\lambda|M_2)\chi_1(\lambda|M_2) = 0 \tag{2.42}$$
$$\chi_0(\lambda|M_3)\chi_1(\lambda|M_3) = 0. \tag{2.43}$$

The above six equations imply that, for each value of λ, one of each pair $\{\chi(0|\lambda, M_i), \chi(1|\lambda, M_i)\}$ must be equal to 0, and the other must be equal to 1.

Now consider the measurement procedure M_* which randomly and with uniform probability chooses an $i \in \{1,2,3\}$, implements measurement M_i, and returns the outcome given by that measurement. In quantum theory, M_* is a POVM represented by

$$\left\{\frac{1}{3}E_{1,0} + \frac{1}{3}E_{2,0} + \frac{1}{3}E_{3,0}, \frac{1}{3}E_{1,1} + \frac{1}{3}E_{2,1} + \frac{1}{3}E_{3,1}\right\} = \left\{\frac{1}{2}\mathbb{I}, \frac{1}{2}\mathbb{I}\right\}. \tag{2.44}$$

The outcomes of M_* are independent of the state being measured, since $\mathrm{Tr}\,\frac{1}{2}\rho\mathbb{I} = \frac{1}{2}$ is independent of ρ. Therefore, M_* is operationally equivalent to a measurement which completely ignores the state of the input system and chooses one of the two outputs uniformly at random. Such a measurement is represented by the set of response functions $\{\frac{1}{2}, \frac{1}{2}\}$. By the assumption of measurement noncontextuality we must have:

$$\{\xi(0|\lambda, M_*), \xi(1|\lambda, M_*)\} = \left\{\frac{1}{2}, \frac{1}{2}\right\}, \tag{2.45}$$

where we have used the outcome-*indeterministic* response functions $\{\xi\left(0|\lambda, M_*\right), \xi\left(1|\lambda, M_*\right)\}$ to represent M_* in the ontological model.

Since M_* is a convex mixture of M_1, M_2, and M_3, we can also represent the response functions of M_* as a convex mixture of the response functions of M_1, M_2, and M_3 (Eq. (2.6)). Thus the following two relations must hold:

$$\frac{1}{3}\chi_0(\lambda|M_1) + \frac{1}{3}\chi_0(\lambda|M_2) + \frac{1}{3}\chi_0(\lambda|M_3) = \frac{1}{2} \tag{2.46}$$

$$\frac{1}{3}\chi_1(\lambda|M_1) + \frac{1}{3}\chi_1(\lambda|M_2) + \frac{1}{3}\chi_1(\lambda|M_3) = \frac{1}{2}. \tag{2.47}$$

However, the constraints implied by Equations (2.38)-(2.40) (i.e. for every λ, the pair of response functions $\{\chi_0(\lambda|M_i), \chi_1(\lambda|M_i)\}$ is equal to either $\{0, 1\}$ or $\{1, 0\}$) contradict Eqs. (2.46) and (2.47). Any choice of assignment of 0 or 1 to each of the terms on the LHS of Eq. (2.46) will give the LHS a sum of 0, $\frac{1}{3}$, $\frac{2}{3}$, or 1, but never $\frac{1}{2}$ (and a similar argument can be made with Eq. (2.47)). Thus we conclude that no generalised noncontextual ontological model can represent the set of three projective-valued measures M_1, M_2, and M_3.

2.4 Local causality as a justification of the assumption of noncontextuality

The assumption of BKS noncontextuality forces the outcome-deterministic ontological model to assign the same response functions to each measurement, regardless of the context in which the observables are measured. A criticism of this assumption, as discussed above, is that measurement of an observable in two different contexts requires two different configurations of an experiment, and it is logically possible that the measurement *is* sensitive to the change of contexts.

Bell's solution to this problem [41, 39] was to restrict the experiment to a more specialised situation, in which the system being measured consists of multiple particles which are simultaneously measured by multiple space-like separated parties. Information cannot travel faster than the speed of light, and thus a change in one party's measurement setting cannot affect the other party's measurement. We will call an ontological model subject to the above constraints (i.e. one in which each party's measurement outcomes only depend on the ontic state of the system and the local choice of measurement settings) a *local ontological model*, (LOM). Bell's theorem [39] states that no LOM can reproduce all the predictions of quantum theory.

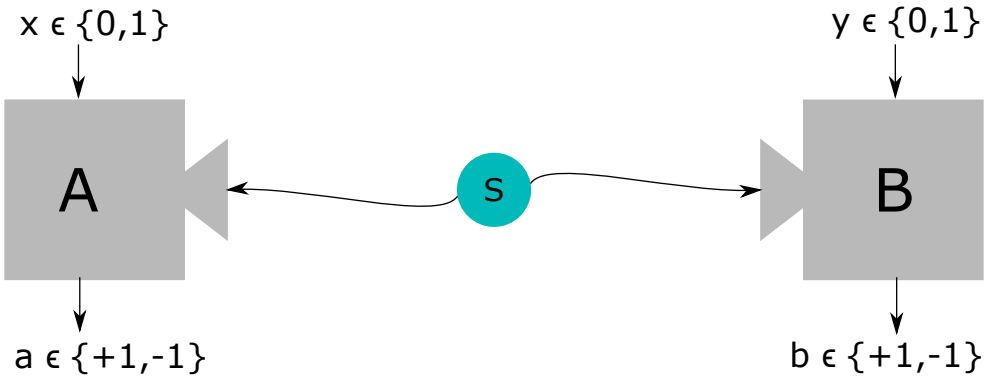

Figure 2.3: The CHSH scenario. A source emits two particles, one travels to Alice, and one travels to Bob. Alice and Bob each can perform one of two measurements on their respective particles, chosen by the random variables x and y, respectively. The outputs of their measurements are either $+1$ or -1. Alice and Bob are space-like separated to ensure that (under the assumption of local causality) Alice's choice of measurement cannot affect Bob's measurement outcome, and *vice versa*.

We will not reproduce Bell's proof of his theorem here, and we will instead present an alternative one found by Clauser, Horne, Shimony, and Holt (CHSH) [48].

2.4.1 The CHSH scenario

The CHSH scenario is illustrated in Figure 2.3. A source emits two particles, one is sent to Alice (labelled with "A") and the other is sent to Bob (labelled with "B").

Alice can perform one of two measurement procedures, M_{A_0} and M_{A_1}, on her particle, and a random variable $x \in \{0, 1\}$ determines which measurement M_{A_x} she performs. After performing her measurement, Alice receives an outcome $a \in \{+1, -1\}$. Similarly, Bob can perform one of the measurements M_{B_0} or M_{B_1} on his particle. A random variable $y \in \{0, 1\}$ determines which measurement M_{B_y} Bob performs, and the measurement returns an outcome $b \in \{+1, -1\}$.

Alice and Bob are space-like separated, and the time interval between Alice's choice of measurement and Bob's implementation of his measurement is less than the time it would take for a light speed signal to travel between Alice and Bob. This ensures that Alice's choice of measurement setting, x, cannot affect Bob's outcome, b. Similarly, Bob's choice of measurement setting, y, is timed such that it cannot affect Alice's outcome, a.

Alice and Bob repeat the experiment on many copies of the system, (randomizing their measurement choices for every copy), and after gathering enough data they share with each other their measurement choices and outcomes for every trial, in order to calculate the expectation values of joint measurements of M_{A_x} and M_{B_y}, defined as:

$$\langle M_{A_x} M_{B_y} \rangle = \sum_{a,b} ab\, p\left(a, b \middle| M_{A_x}, M_{B_y}\right), \tag{2.48}$$

where $p\left(a, b \middle| M_{A_x}, M_{B_y}\right)$ is the probability of obtaining outcomes a and b given measurements procedures M_{A_x} and M_{B_y} were performed.

The CHSH parameter, S_{CHSH}, is a linear combination of the four possible expectation values, defined as:

$$S_{\text{CHSH}} = \langle M_{A_0} M_{B_0} \rangle + \langle M_{A_0} M_{B_1} \rangle + \langle M_{A_1} M_{B_0} \rangle - \langle M_{A_1} M_{B_1} \rangle. \tag{2.49}$$

The CHSH inequality states that, for any local ontological model, the following inequality holds

$$|S_{\text{CHSH}}| \underset{\text{LOM}}{\leq} 2. \tag{2.50}$$

However, quantum theory can violate this inequality, and specifically

$$|S_{\text{CHSH}}| \underset{\text{QT}}{\leq} 2\sqrt{2}. \tag{2.51}$$

2.4.2 Proof of the CHSH inequality

In a general (i.e. not necessarily local) ontological model of the CHSH scenario, the source enacts some preparation procedure P which prepares the two-particle system in the state λ with some probability $\mu(\lambda|P)$. The expectation values $\langle M_{A_x} M_{B_y} \rangle$ are calculated with:

$$\langle M_{A_x} M_{B_y} \rangle = \sum_{a,b} ab \int d\lambda \mu(\lambda|P) \xi\left(a, b|\lambda, x, y\right). \tag{2.52}$$

Since Alice and Bob are space-like separated, Alice's outcome cannot depend on Bob's choice of measurement setting, nor should it depend on Bob's measurement outcome. Similarly, Bob's outcome cannot depend on Alice's measurement setting or outcome. Hence, for a LOM, the measurement response function can be factored in the following way[3].

$$\xi\left(a, b|\lambda, x, y\right) = \xi\left(a|\lambda, x\right) \xi\left(b|\lambda, y\right). \tag{2.53}$$

[3]We can interpret this factorization of the joint measurement response function as an application of the assumption of measurement noncontextuality. The measurement M_{A_x} is measured in two contexts: one

Restricting our attention to one specific ontic state λ, the expectation value of the joint measurement of M_{A_x} and M_{B_y} is:

$$\langle M_{A_x|\lambda} M_{B_y|\lambda} \rangle = \sum_{a,b} ab\, \xi\left(a|\lambda, x\right) \xi\left(b|\lambda, y\right), \tag{2.54}$$

which is equal to:

$$\left\langle M_{A_x|\lambda} M_{B_y|\lambda} \right\rangle = \sum_a a\xi(a|\lambda, x) \sum_b b\xi(b|\lambda, y) \tag{2.55}$$

$$= \left\langle M_{A_x|\lambda} \right\rangle \left\langle M_{B_y|\lambda} \right\rangle. \tag{2.56}$$

The CHSH parameter for a specific value of λ is:

$$S_{\text{CHSH}|\lambda} = \left\langle M_{A_0|\lambda} M_{B_0|\lambda} \right\rangle + \left\langle M_{A_0|\lambda} M_{B_1|\lambda} \right\rangle + \left\langle M_{A_1|\lambda} M_{B_0|\lambda} \right\rangle - \left\langle M_{A_1|\lambda} M_{B_1|\lambda} \right\rangle \tag{2.57}$$

which, with (2.56), can be simplified to:

$$S_{\text{CHSH}|\lambda} = \left\langle M_{A_0|\lambda} \right\rangle \left\langle M_{B_0|\lambda} \right\rangle + \left\langle M_{A_0|\lambda} \right\rangle \left\langle M_{B_1|\lambda} \right\rangle + \left\langle M_{A_1|\lambda} \right\rangle \left\langle M_{B_0|\lambda} \right\rangle - \left\langle M_{A_1|\lambda} \right\rangle \left\langle M_{B_1|\lambda} \right\rangle \tag{2.58}$$

$$= \left\langle M_{A_0|\lambda} \right\rangle \left(\left\langle M_{B_0|\lambda} \right\rangle + \left\langle M_{B_1|\lambda} \right\rangle \right) + \left\langle M_{A_1|\lambda} \right\rangle \left(\left\langle M_{B_0|\lambda} \right\rangle - \left\langle M_{B_1|\lambda} \right\rangle \right) \tag{2.59}$$

Taking the absolute value of both sides, and using the fact that the single-particle expectation values $\left\langle M_{A_x|\lambda} \right\rangle$ and $\left\langle M_{B_y|\lambda} \right\rangle$ must both lie in the interval $[-1, 1]$, we have:

$$\left| S_{\text{CHSH}|\lambda} \right| \leq \left| \left\langle M_{A_0|\lambda} \right\rangle \left(\left\langle M_{B_0|\lambda} \right\rangle + \left\langle M_{B_1|\lambda} \right\rangle \right) \right| + \left| \left\langle M_{A_1|\lambda} \right\rangle \left(\left\langle M_{B_0|\lambda} \right\rangle - \left\langle M_{B_1|\lambda} \right\rangle \right) \right| \tag{2.60}$$

$$\leq \left| \left\langle M_{B_0|\lambda} \right\rangle + \left\langle M_{B_1|\lambda} \right\rangle \right| + \left| \left\langle M_{B_0|\lambda} \right\rangle - \left\langle M_{B_1|\lambda} \right\rangle \right| \tag{2.61}$$

$$\leq 2. \tag{2.62}$$

Finally, we average over all λ:

$$\left| S_{\text{CHSH}} \right| = \left| \int d\lambda \mu(\lambda|P) S_{\text{CHSH}|\lambda} \right| \tag{2.63}$$

$$\leq \int d\lambda \mu(\lambda|P) \left| S_{\text{CHSH}|\lambda} \right|, \tag{2.64}$$

and we arrive at the result:

$$\left| S_{\text{CHSH}} \right| \underset{\text{LOM}}{\leq} 2. \tag{2.65}$$

context is as part of a joint measurement with M_{B_0}, and the other context is as part of a joint measurement with M_{B_1}. Because of the space-like separation of Alice and Bob, the measurement that Bob chooses to perform can not in any way affect Alice's measurement outcomes. When the factorization in Eq. (2.53) is performed, we are assuming that since Alice's outcome statistics for measurement M_{A_x} are independent of the measurement context we can represent M_{A_x} with the same response functions in each context; this is precisely the assumption of measurement noncontextuality.

2.4.3 Quantum violation of CHSH inequality

With the proper choice of both a two-particle quantum state and pairs of measurements for both Alice and Bob, quantum theory predicts that the CHSH inequality can be violated. Tsirelson showed that the maximum CHSH parameter achievable by quantum theory is $2\sqrt{2}$ [49].

One way to achieve this maximum violation is with a source that emits the singlet state $|\Psi^-\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A |1\rangle_B - |1\rangle_A |0\rangle_B)$, where the subscripts A and B label the particle that is sent to Alice and Bob, respectively. Letting $|\theta\rangle = \cos\theta/2\,|0\rangle + \sin\theta/2\,|1\rangle$ and $\left|\theta^\perp\right\rangle = |\theta + \pi\rangle$, we can define a projective measurement observable as $O_\theta = |\theta\rangle\langle\theta| - \left|\theta^\perp\right\rangle\left\langle\theta^\perp\right|$.

The expectation value of the joint measurement $O_{\theta_x} O_{\theta_y}$ on the state $|\Psi^-\rangle$ is $\left\langle O_{\theta_x} O_{\theta_y}\right\rangle = -\cos(\theta_x - \theta_y)$. If Alice chooses $\theta_0 = 0$ and $\theta_1 = \pi/2$ for her two measurements, and if Bob chooses $\theta_0 = \pi/4$ and $\theta_1 = -\pi/4$ for his two measurements, then Alice and Bob will measure $S_{\text{CHSH}} = -2\sqrt{2}$, violating inequality (2.65).

Chapter 3

The framework of generalised probabilistic theories

Notes and acknowledgements

This chapter contains material from Section II of work that has appeared on the arXiv as [1]:

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," *arXiv:1710.05948*, (2017).

Author contributions

M. D. Mazurek and **R. W. Spekkens** wrote the first draft of this part of the work.

All authors contributed to the final draft.

3.1 Chapter introduction

So far in this thesis we have explored the framework of quantum theory as well as the ontological model framework. In this chapter we will introduce the framework of *generalised probabilistic theories* (GPTs). The GPT framework is operationally motivated, and it

describes system preparation and measurement procedures exclusively in terms of possible measurement-outcome probabilities. As we will see, the GPT framework is extremely general, and it can be used to model a large class of theories including, for example, classical theory, quantum theory, and even so-called "super"-quantum theories, which can exhibit stronger-than-quantum correlations between pairs of space-like separated particles.

A recent application of the GPT framework in the field of quantum foundations has been as a tool for deriving axiomatizations of quantum theory [50, 51, 52, 53], the method of such axiomatizations being to begin with a list of desirable physical principles, then derive the class of GPTs consistent with these principles, and finally compare this class to quantum theory.

The GPT framework is also useful for analysing experimental data (as is done in the experiments presented in Chapters 4 and 5). By analysing data within the GPT framework, as opposed to within a specific physical theory such as quantum theory, the conclusions drawn from the data can be very general, and can be used to compare the accuracies of competing physical theories.

In this chapter we will introduce the GPT framework with the application of data analysis in mind, which will provide the background necessary to understand the later chapters of this thesis.

3.2 Basics

For any system, in any physical theory, there will in general be many possible ways for it to be prepared, transformed, and measured. Here, each preparation procedure, transformation procedure and measurement procedure is conceived as a list of instructions for what to do in the laboratory. The different combinations of possibilities for each procedure defines a collection of possible experimental configurations. We will here restrict our attention to experimental configurations of the prepare-and-measure variety: these are the configurations where there is no transformation intervening between the preparation and the measurement and where the measurement is terminal (which is to say that the system does not persist after the measurement). We further restrict our attention to binary-outcome measurements.

A GPT aims to describe only the operational phenomenology of a given experiment. In the case of a prepare-and-measure experiment, it aims to describe only the relative probabilities of the different outcomes of each possible measurement procedure when it

is implemented following each possible preparation procedure. For binary-outcome measurements, it suffices to specify the probability of one of the outcomes since the other is determined by normalization. If we denote the outcome set by $\{0, 1\}$, then it suffices to specify the probability of the event of obtaining outcome 0 in measurement M. This event will be termed an *effect* and denoted $[0|M]$.

Thus a GPT specifies a probability $p(0|P, M)$ for each preparation P and measurement M. Denoting the cardinality of the set of all preparations (respectively all measurements) by m (respectively n), the set of these probabilities can be organized into an $m \times n$ matrix, denoted D, where the rows correspond to distinct preparations and the columns correspond to distinct effects,

$$D \equiv \begin{pmatrix} p(0|P_1, M_1) & p(0|P_1, M_2) & \cdots & p(0|P_1, M_n) \\ p(0|P_2, M_1) & p(0|P_2, M_2) & \cdots & p(0|P_2, M_n) \\ \cdots & \cdots & \cdots & \\ p(0|P_m, M_1) & p(0|P_m, M_2) & \cdots & p(0|P_m, M_n) \end{pmatrix} \tag{3.1}$$

We refer to D as the *probability matrix* associated to the physical theory. Because it specifies the probabilities for *all* possibilities for the preparations and the measurements, it contains all of the information about the putative physical theory for prepare-and-measure experiments.[1]

Defining

$$k \equiv \text{rank}(D)$$

then one can factor D into a product of two rectangular matrices,

$$D = SE \tag{3.2}$$

where S is an $(m \times k)$ matrix and E is a $(k \times n)$ matrix.

Denoting the ith row of S by the row vector $\mathbf{s}_{P_i}^T$ (where T denotes transpose) and the jth column of E by the column vector $\mathbf{e}_{[0|M_j]}$, we can write

$$D = \begin{pmatrix} \mathbf{s}_{P_1}^T \\ \mathbf{s}_{P_2}^T \\ \cdots \\ \mathbf{s}_{P_m}^T \end{pmatrix} \begin{pmatrix} \mathbf{e}_{[0|M_1]} & \mathbf{e}_{[0|M_2]} & \cdots & \mathbf{e}_{[0|M_n]} \end{pmatrix}, \tag{3.3}$$

[1]Note that although the presentation as a table suggests that the sets of preparations and measurements are discrete, there could in fact be a continuum of possibilities for each set. If the continuous variable labelling the preparations in the theory is x and that labelling the measurements in the theory is y, then the complete information about the physical theory is given by the function $f(x, y) := p(0|P_x, M_y)$. The GPT is a theoretical abstraction, so that it is acceptable if it is presumed to contain such continua.

so that

$$p(0|P_i, M_j) = \mathbf{s}_{P_i} \cdot \mathbf{e}_{[0|M_j]}. \tag{3.4}$$

Factoring D in this way allows us to associate with each preparation P a k-dimensional vector $\mathbf{s}_P$ and with each effect $[0|M]$ a k-dimensional vector $\mathbf{e}_{[0|M]}$ such that the probability of obtaining the effect $[0|M]$ on the preparation P is recovered as their inner product, $p(0|P, M) = \mathbf{s}_P \cdot \mathbf{e}_{[0|M]}$. The vectors $\mathbf{s}_P$ and $\mathbf{e}_{[0|M]}$ will be termed *GPT state vectors* and *GPT effect vectors* respectively. A particular GPT is specified by the sets of all allowed GPT state and effect vectors, denoted by $\mathcal{S}$ and $\mathcal{E}$, respectively.

Because the n GPT effect vectors associated to the set of all measurement effects lie in a k-dimensional vector space, only k of them are linearly independent. Any set of k measurement effects whose associated GPT effect vectors form a basis for the space will be termed a *tomographically complete* set of measurement effects. The terminology stems from the fact that if one seeks to deduce the GPT state vector of an unknown preparation from the probabilities it assigns to a set of characterized measurement effects (the GPT analogue of quantum state tomography) then this set of GPT effect vectors must form a basis of the k-dimensional space. Similarly, any set of k preparations whose associated GPT state vectors form a basis for the space will be termed *tomographically complete* because to deduce the GPT effect vector of an unknown measurement effect from the probabilities assigned to it by a set of known preparations, the GPT state vectors associated to the latter must form a basis. For any GPT, we necessarily have that the rank of D satisfies $k \leq \min\{m, n\}$, but in general, we expect k to be much smaller than m or n.

There is a freedom in the decomposition of Eq. (3.2). Specifically, for any invertible $(k \times k)$ matrix R, we have $D = SE = (SR^{-1})(RE)$. Thus, there are many decompositions of D of the type described. The vectors $\{\mathbf{s}_{P_i}\}_i$ and $\{\mathbf{e}_{[0|M_j]}\}_j$ depend on the specific decomposition chosen. However, for any two choices of decompositions SE and $S'E'$, the vectors $\{\mathbf{s}_{P_i}\}_i$ and $\{\mathbf{s}'_{P_i}\}_i$ (and the vectors $\{\mathbf{e}_{[0|M_j]}\}_j$ and $\{\mathbf{e}'_{[0|M_j]}\}_j$) are always related by a linear transformation. Note that any basis of the k-dimensional vector space remains so under a linear transformation, so the property of being tomographically complete is independent of the choice of representation.

It is worth noting that for *any* physical theory, the GPT framework provides a complete description of its operational predictions for prepare-and-measure experiments. In this sense, the GPT framework is completely general. Furthermore, one can show that under a very weak assumption it provides the most efficient description of the theory, in the sense that it is a description with the smallest number of parameters. The weak assumption is that it is possible to implement arbitrary convex mixtures of preparations without altering the functioning of each preparation in the mixture, so that for any set of GPT state vectors

that are admitted in the theory, all of the vectors in their convex hull are also admitted in the theory. See Theorem 1 of Ref. [54] for the proof.

We will here make this weak assumption and restrict our attention to GPTs wherein any convex mixture of preparation procedures is another valid preparation procedure, so that the set of GPT state vectors is convex [50]. In this case, we can refer to the set $\mathcal{S}$ of GPT states in a theory as its *GPT state space*. We also make the weak assumption that any convex mixture of measurements and any classical post-processing of a measurement is another valid measurement. This implies that the set of GPT effect vectors lie in the intersection of two cones: the one defined by taking the convex hull of all positive multiples of the GPT effect vectors, and the one defined by demanding positivity of the set of complementary effects (for a given effect $\mathbf{e}$, its complementary effect $\bar{\mathbf{e}}$ is the one for which, the sum with $\mathbf{e}$ gives the unit effect, defined below), a shape that we will refer to as a "diamond". In this case, we can refer to the set $\mathcal{E}$ of GPT effects in a theory as its *GPT effect space*.

It is worth noting that although GPTs which fail to be closed under convex mixtures and classical post-processing are of theoretical interest — there are interesting foils to quantum theory of this type [38, 55] — one does not expect them to be candidates for the true GPT describing nature because there seems to be no obstacle in practice to mixing or post-processing procedures in an arbitrary way. To put it another way, the evidence suggests that the GPT describing nature must include classical probability theory as a subtheory, thereby providing the resources for implementing arbitrary mixtures and post-processings.

Distinct physical theories (i.e., distinct GPTs) are distinguished by the *shapes* of the GPT state space and the GPT effect space, where these shapes are defined up to a linear transformation, as described earlier.

We end by highlighting some conventions we adopt in representing GPTs. Define the "unit" measurement effect as the one which occurs with probability 1 for all preparations (it is represented by a column of 1s in D), and denote it by $\mathbf{u}$. Because each $\mathbf{s}_P$ will have an inner product of 1 with $\mathbf{u}$ (by normalization of probability), it follows that there are only $k-1$ free parameters in the GPT state vector. We make a conventional choice (i.e., a particular choice within the freedom of linear transformations) to represent the unit effect by the GPT effect vector $(1, 0, 0, \dots)^T$. This choice forces the first component of all of the GPT state vectors to be 1. In this case, one can restrict the search for factorizations $D = SE$ to those for which the first column of S is a column of 1s. It also follows that the projection of all GPT state vectors along one of the axes of the k-dimensional vector space has value 1, and consequently it is useful to only depict the projection of the GPT state vectors into the complementary $(k-1)$-dimensional subspace.

3.3 GPT representations of some example physical theories

In this section we will illustrate and clarify the ideas presented above with a few helpful examples.

Qubit

First, consider a 2-level quantum system (i.e., a qubit). As we saw in Section 1.1.4, a qubit state ρ is defined as $\rho = \frac{1}{2}\mathbf{s}\cdot\boldsymbol{\sigma}$ (for $\mathbf{s} = \begin{pmatrix} s_0 & s_1 & s_2 & s_3 \end{pmatrix}$) a measurement effect operator Q is defined as $Q = \mathbf{e}\cdot\boldsymbol{\sigma}$ (for $\mathbf{e} = \begin{pmatrix} e_0 & e_1 & e_2 & e_3 \end{pmatrix}$), and $\mathrm{Tr}(\rho E) = \mathbf{r}\cdot\mathbf{e}$. Thus the allowed values of $\mathbf{s}$ define the qubit GPT state space, and the allowed values of $\mathbf{e}$ define the qubit GPT effect space. Hence, the Bloch representation of a qubit that we presented in Section 1.1.4 is a valid GPT representation!

As noted above, this geometric representation of the quantum state and effect spaces is only one possibility among many. If we define a linear transformation of the state space by any invertible 4×4 matrix and we take the corresponding inverse linear transformation on the effect space, the new state and effect spaces will also provide an adequate representation of all prepare-and-measure experiments on a single qubit. (Note that implementing a linear transformation of this form is equivalent to representing quantum states and effects with respect to a different basis of Hermitian operators.)

Classical bit

Classical probabilistic theories can also be formulated within the GPT framework. Consider the simplest case of a classical system with two possible physical states, i.e., a classical bit, for which $k = 2$. The set of possible preparations of this system is simply the set of normalized probability distributions on a bit, $\vec{\mu} = (\mu_0, \mu_1)$, where $0 \leq \mu_0, \mu_1 \leq 1$ and $\mu_0 + \mu_1 = 1$. The most general measurement effect is a pair of probabilities, specifying the probability of that effect occuring for each value of the bit, that is, $\vec{\xi} = (\xi_0, \xi_1)$ where $0 \leq \xi_0, \xi_1 \leq 1$. The probability of a particular measurement effect occuring when implemented on a particular preparation is clearly just the inner product of these, $\vec{\mu}\cdot\vec{\xi}$. The positivity and normalization constraints imply that the convex set of state vectors describes a line segment from $(1, 0)$ to $(0, 1)$, and the set of effect vectors is the square region with vertices $(0, 0), (1, 0), (0, 1)$ and $(1, 1)$.

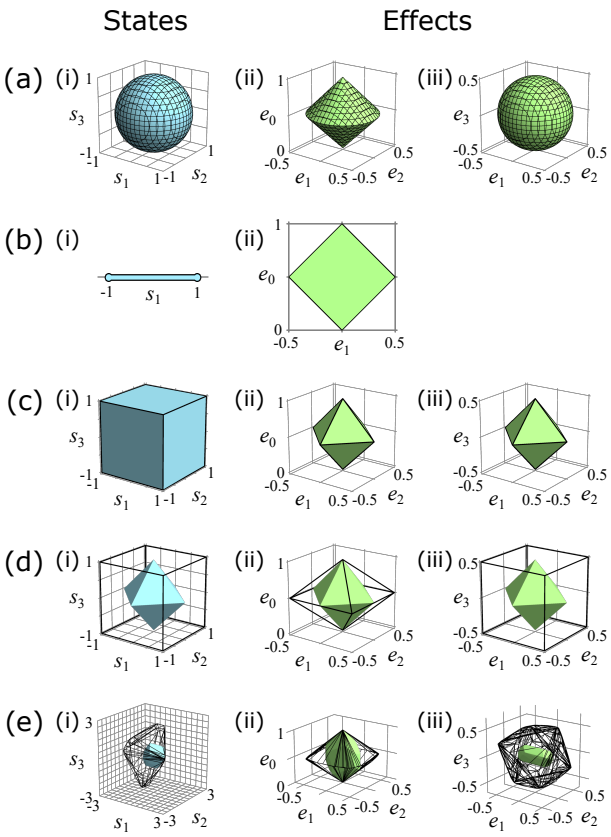

Figure 3.1: Some paradigm examples of GPTs. The solid shapes represent the true state and effect spaces for that GPT, while the black wireframe shapes represent the duals of these (for the duality relation described in Sec. 3.4). (i) The true state space (solid blue) and its dual (wireframe). (ii)-(iii) The true effect space (solid green) and its dual (wireframe). For the cases where $k = 4$, the effect spaces are 4d, and we depict them by a pair of 3d projections. (a) A qubit ($k = 4$). (b) A classical bit ($k = 2$). (c) The $k = 4$ system in Boxworld. (c) The convex closure of the Spekkens toy theory for the simplest system ($k = 4$). (e) A generic GPT with $k = 4$, obtained from a randomly generated rank-4 matrix of probabilities.

For ease of comparison with our other examples of GPTs, it is useful to consider a linear transformation of this representation, corresponding geometrically to a rotation by 45 degrees. We represent each preparation by a state vector $\mathbf{s} = (1, s_1)$, where $-1 \leq s_1 \leq 1$, and each measurement effect by an effect vector $\mathbf{e} = (e_0, e_1)$ where $-1/2 \leq e_1 \leq 1/2$ and $e_0 \geq |e_1|$ and $e_0 \leq 1 - |e_1|$ (with the experimental probabilities still given by their inner product, $\mathbf{s} \cdot \mathbf{e}$). The convex set of these state vectors can then be depicted as a horizontal line segment, and the set of effect vectors by a diamond with a line segment at its base, as in Fig. 3.1(b). This representation makes it clear that the state and effect spaces of a classical bit are contained within those of a qubit (as the quantum states and effects whose representation as operators are diagonal in some fixed basis of the Hilbert space).

Generalised no-signalling theory

One can also consider GPTs that are neither classical nor quantum. In the GPT known as "Boxworld" [51, 56] (originally called "generalised no-signalling theory"), correlations can be stronger than in quantum theory, violating Bell inequalities by an amount in excess of the maximum quantum violation. The $k = 3$ system in Boxworld, known as the "generalized no-signalling bit", has received a great deal of attention. A pair of such systems can generate the stronger-than-quantum correlations known as a Popescu-Rohrlich box [57] from which the name Boxworld derives. These achieve a CHSH inequality violation equal to the algebraic maximum. Such correlations are achievable in Boxworld because there are some states that respond deterministically to multiple effects, and there are also some effects that respond deterministically to multiple states. Boxworld also has a $k = 4$ system, which shares features of the generalized no-signalling bit and is, in certain respects, more straightforward to compare to a qubit. It is the latter that we depict in Fig. 3.1(c).

Spekkens toy theory

Another alternative to classical and quantum theories is the toy theory introduced by Spekkens [58]. We here consider a variant of this theory, wherein one closes under convex combinations. The simplest system has $k = 4$ and has the state and effect spaces depicted in Figure 3.1(d). These state and effect spaces are strictly contained within those of the classical theory for a system with four physical states (the $k = 4$ system in the classical theory), which corresponds to the fact that the theory can be understood as the result of imposing an additional restriction relative to what can be achieved classically.

Randomly generated theory

Finally, Fig. 3.1(e) illustrates a generic example of a GPT with $k = 4$. We constructed this GPT by generating a rank 4 matrix of random probability data. We defined the convex closure of the set of GPT states and measurement effects represented by this data as the full state and effect spaces of a physical theory, which gives us our randomly generated theory.

3.4 Dual spaces

Finally, we review the notion of *dual* GPT state and effect spaces. We will call a vector $\mathbf{s} \in \mathbb{R}^k$ a *logically possible state* if it assigns a valid probability to every measurement effect allowed by the GPT. Mathematically, the space of logically possible states, denoted $\mathcal{S}_{\text{logical}}$, contains all $\mathbf{s} \in \mathbb{R}^k$ such that $\forall \mathbf{e} \in \mathcal{E} : 0 \leq \mathbf{s} \cdot \mathbf{e} \leq 1$ and such that $\mathbf{s} \cdot \mathbf{u} = 1$. From this definition, it is clear that $\mathcal{S}_{\text{logical}}$ is the intersection of the geometric dual of $\mathcal{E}$ and the hyperplane defined by $\mathbf{s} \cdot \mathbf{u} = 1$; as a shorthand, we will refer to $\mathcal{S}_{\text{logical}}$ simply as "the dual of $\mathcal{E}$", and denote the relation by $\mathcal{S}_{\text{logical}} \equiv \text{dual}(\mathcal{E})$. Analogously, the set of logically possible effects, denoted $\mathcal{E}_{\text{logical}}$, contains all $\mathbf{e} \in \mathbb{R}^k$ such that $\forall \mathbf{s} \in \mathcal{S} : 0 \leq \mathbf{s} \cdot \mathbf{e} \leq 1$. Defining the set of subnormalized states by $\hat{\mathcal{S}} \equiv \{w\mathbf{s} : \mathbf{s} \in \mathcal{S}, w \in [0, 1]\}$, $\mathcal{E}_{\text{logical}}$ is the geometric dual of $\hat{\mathcal{S}}$. For simplicity, we will refer to $\mathcal{E}_{\text{logical}}$ simply as "the dual of $\mathcal{S}$", and denote the relation by $\mathcal{E}_{\text{logical}} \equiv \text{dual}(\mathcal{S})$.

GPTs in which $\mathcal{S}_{\text{logical}} = \mathcal{S}$ and $\mathcal{E}_{\text{logical}} = \mathcal{E}$ (the two conditions are equivalent) are said to satisfy the *no-restriction hypothesis* [52]. In a theory that satisfies the no-restriction hypothesis, every logically allowed GPT effect vector corresponds to a physically allowed measurement, and (equivalently) every logically allowed GPT state vector corresponds to a physically allowed preparation. In theories wherein $\mathcal{S}_{\text{logical}} \neq \mathcal{S}$ and $\mathcal{E}_{\text{logical}} \neq \mathcal{E}$, by contrast, there are vectors that do not correspond to physically allowed states but nonetheless assign valid probabilities to all physically allowed effects, and there are vectors that do not correspond to physically allowed effects but are nonetheless assigned valid probabilities by all physically allowed states.

For each of the examples in Fig. 3.1, we have depicted the dual to the effect space alongside the state space and the dual of the state space alongside the effect space, as wireframes. Quantum theory, classical probability theory, and Boxworld provide examples of GPTs that satisfy the no-restriction hypothesis, as illustrated in Fig. 3.1(a),(b),(c), while the GPTs presented in Fig. 3.1(d),(e) are examples of GPTs that violate it.

Chapter 4

An experimental test of noncontextuality without unphysical idealizations

Notes and acknowledgements

This chapter contains work that has been published as [2]:

M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch and R. W. Spekkens, "An experimental test of noncontextuality without unphysical idealizations" *Nature Communications*, **7**:11780 (2016).

Author contributions

Ravi Kunjwal and **Rob Spekkens** proved the noncontextuality inequality.

Michael Mazurek and **Kevin Resch** designed the experiment and implemented the numerical analysis of the data.

Michael Mazurek performed the experiment and showed that the bound of the noncontextuality inequality was tight.

Matthew Pusey developed the method to overcome the problem of inexact operational equivalence.

All authors contributed to writing the manuscript.

4.1 Overview

To make precise the sense in which nature fails to respect classical physics, one requires a formal notion of classicality. Ideally, such a notion should be defined operationally, so that it can be subject to direct experimental test, and it should be applicable in a wide variety of experimental scenarios so that it can cover the breadth of phenomena thought to defy classical understanding. Bell's notion of local causality fulfills the first criterion but not the second. The notion of noncontextuality fulfills the second criterion, but it is a long-standing question whether it can be made to fulfill the first. Previous attempts to test noncontextuality have all assumed idealizations that real experiments cannot achieve, namely noiseless measurements and exact operational equivalences. Here we show how to devise tests that are free of these idealizations. We perform a photonic implementation of one such test, ruling out noncontextual models with high confidence.

4.2 Introduction

Making precise the manner in which a quantum world differs from a classical one is a surprisingly difficult task. The most successful attempt, due to Bell [39], shows a conflict between quantum theory and a feature of classical relativistic theories termed *local causality*, which asserts that no causal influences propagate faster than light. But the latter assumption can only be tested for scenarios wherein there are two or more systems that are space-like separated. And yet few believe that this highly specialized situation is the only point where the quantum departs from the classical. A leading candidate for a notion of nonclassicality with a broader scope is the failure of quantum theory to admit of a noncontextual model, as proven by Kochen and Specker [40]. Recent work has highlighted how this notion lies at the heart of many phenomena that are taken to be distinctly quantum: the fact that quasi-probability representations go negative [59, 60], the existence of quantum advantages for cryptography [61] and for computation [62, 63, 64], and the possibility of anomalous weak values [65]. Consequently, the study of noncontextuality has not only foundational significance but practical applications as well.

An experimental refutation of noncontextuality would demonstrate that the conflict with noncontextual models is not only a feature of quantum theory, but of nature itself, and hence also of any successor to quantum theory. The requirements for such an experimental test, however, have been a subject of much controversy [66, 67, 68, 69, 70, 71, 72].

A fundamental problem with most proposals for testing noncontextuality [73, 74, 75, 76, 47, 77, 78, 79], and experiments performed to date [80, 81, 82, 83, 84, 85, 86, 87],

is that they seek to test a notion of noncontextuality which posits that measurements have a deterministic response in the noncontextual model. It has been shown that such determinism is only justified under the idealization that measurements are noiseless [88], which is never satisfied precisely by any real experiment. We refer to this issue as the *problem of noisy measurements*.

Another critical problem with previous proposals is the fact that the assumption of noncontextuality can only be brought to bear when two measurement events (an event is a measurement and an outcome) are *operationally equivalent*, which occurs when the two events are assigned exactly the same probability by all preparation procedures [38]; in this case they are said to differ only by the measurement *context*. In a real experiment, however, one never achieves the ideal of precise operational equivalence. Previous work on testing noncontextuality—including the only experiment to have circumvented the problem of noisy measurements (by focusing on preparations) [61]—has failed to provide a satisfactory account of how the deviation from strict operational equivalence should be accounted for in the interpretation of the results. We term this problem the *problem of inexact operational equivalence*.

In this work, we solve both of the above problems. We contend with the problem of noisy measurements by devising a test of a generalised notion of noncontextuality, proposed in Reference [38], that allows general measurements to have an *indeterministic* response while reducing to the traditional notion in the idealized case of projective quantum measurements. For the problem of inexact operational equivalence, whereas some have been led to consider *modifying* the definition of noncontextuality so that it applies to pairs of procedures that are merely *close* to operationally equivalent [89, 90], we circumvent the problem by demonstrating a general technique that appeals to equivalences not among the procedures themselves, but certain convex mixtures thereof. Of course, any judgment of operational equivalence of measurements (preparations) rests on an assumption about which sets of preparations (measurements) are sufficient to establish such equivalence, that is, which sets are *tomographically complete*. We here assume that the cardinality of a tomographically-complete set of measurements (preparations) for a photon's polarization is three (four), as it is in quantum theory. We collect some experimental evidence for this assumption—another improvement over previous experiments—but the possibility of its failure is the most significant remaining loophole for tests of noncontextuality. For Bell's notion of local causality, the theoretical work of Clauser *et al.* [48] was critical to enabling an experimental test without unwarranted idealizations, e.g., without the perfect anti-correlations presumed in Bell's original proof [39]. Similarly, the theoretical innovations we introduce here make it possible for the first time to subject noncontextuality to an experimental test without the idealizations described above. We report on a quantum-

optical experiment of this kind, the results of which rule out noncontextual models with high confidence.

4.3 Results

4.3.1 A noncontexuality inequality

According to the operational approach proposed in Ref. [38], to assume noncontextuality is to assume a constraint on model-construction, namely, that *if procedures are statistically equivalent at the operational level then they ought to be statistically equivalent in the underlying model.*

Operationally, a system is associated with a set $\mathcal{M}$ (resp. $\mathcal{P}$) of physically possible measurement (resp. preparation) procedures. An *operational theory* specifies the possibilities for the conditional probabilities $\{p(X|P,M) : P \in \mathcal{P}, M \in \mathcal{M}\}$ where X ranges over the outcomes of measurement M. In an *ontological model* of such a theory, the causal influence of the preparation on the measurement outcome is mediated by the *ontic state* of the system, that is, a full specification of the system's physical properties. We denote the space of ontic states by Λ. It is presumed that when the preparation P is implemented, the ontic state of the system, $\lambda \in \Lambda$, is sampled from a probability distribution $\mu(\lambda|P)$, and when the system is subjected to the measurement M, the outcome X is distributed as $\xi(X|M,\lambda)$. Finally, for the model to reproduce the experimental statistics, we require that

$$\sum_{\lambda\in\Lambda} \xi(X|M,\lambda)\mu(\lambda|P) = p(X|M,P). \tag{4.1}$$

Suppose there is a measurement procedure, M_*, that is operationally indistinguishable from a fair coin flip: it always gives a uniformly random outcome regardless of the preparation procedure,

$$p(X=0,1|M_*,P) = \frac{1}{2},\ \forall P \in \mathcal{P}. \tag{4.2}$$

In this case, noncontextuality (see Section 2.3.1) dictates that in the underlying model, the measurement should also give a uniformly random outcome regardless of the ontic state of the system,

$$\xi(X=0,1|M_*,\lambda) = \frac{1}{2},\ \forall \lambda \in \Lambda. \tag{4.3}$$

In other words, because M_* appears operationally to be just like a coin flip, noncontextuality dictates that physically it must be just like a coin flip.

The second application of noncontextuality is essentially a time-reversed version of the first. Suppose there is a triple of preparation procedures, P_1, P_2 and P_3, that are operationally indistinguishable from one another: no measurement reveals any information about which of these preparations was implemented,

$$\forall M \in \mathcal{M} : p(X|M, P_1) = p(X|M, P_2) = p(X|M, P_3). \tag{4.4}$$

In this case, noncontextuality dictates that in the underlying model, the ontic state of the system does not contain any information about which of these preparation procedures was implemented,

$$\forall \lambda \in \Lambda : \mu(\lambda|P_1) = \mu(\lambda|P_2) = \mu(\lambda|P_3). \tag{4.5}$$

In other words, because it is impossible, operationally, to extract such information, noncontextuality dictates that physically, the information is not present in the system.

Suppose that M_* can be realized as a uniform mixture of three other binary-outcome measurements, denoted M_1, M_2 and M_3. That is, one implements M_* by uniformly sampling $t \in \{1, 2, 3\}$, implementing M_t, then outputting its outcome as the outcome of M_* (ignoring t thereafter). Finally, suppose that each preparation P_t can be realized as the equal mixture of two other preparation procedures, denoted $P_{t,0}$ and $P_{t,1}$.

Consider implementing M_t on $P_{t,b}$, and consider the average *degree of correlation* between the measurement outcome X and the preparation variable b:

$$A \equiv \frac{1}{6} \sum_{t \in \{1,2,3\}} \sum_{b \in \{0,1\}} p(X = b|M_t, P_{t,b}). \tag{4.6}$$

We now show that noncontextuality implies a nontrivial bound on A.

The proof is by contradiction. In order to have perfect correlation on average, we require perfect correlation in each term, which implies that for all ontic states λ assigned nonzero probability by $P_{t,b}$, the measurement M_t must respond deterministically with the $X = b$ outcome. Given that P_t is an equal mixture of $P_{t,0}$ and $P_{t,1}$, it follows that for all ontic states λ assigned nonzero probability by P_t, the measurement M_t must have a deterministic response, i.e., $\xi(X = b|M_t, \lambda) \in \{0, 1\}$.

But Equation (4.5) (which follows from the assumption of noncontextuality) asserts that the preparations P_1, P_2 and P_3 must assign nonzero probability to precisely the *same*

set of ontic states. Therefore, to achieve perfect correlation on average, each measurement must respond deterministically to *all* the ontic states in this set.

Now note that by the definition of M_*, the probability of its outcome $X = b$ is $\xi(X = b|M_*, \lambda) = \frac{1}{3}\sum_{t\in\{1,2,3\}} \xi(X = b|M_t, \lambda)$. But then Equation (4.3) (which follows from the assumption of noncontextuality) says

$$\frac{1}{3}\sum_{t\in\{1,2,3\}} \xi(X = b|M_t, \lambda) = \frac{1}{2}. \tag{4.7}$$

For each deterministic assignment of values, $(\xi(X = b|M_1, \lambda), \xi(X = b|M_2, \lambda), \xi(X = b|M_3, \lambda)) \in \{(0, 0, 0), (0, 0, 1), \ldots, (1, 1, 1)\}$, the constraint of Equation (4.7) is violated. It follows, therefore, that for a given λ, one of M_1, M_2 or M_3 must fail to have a deterministic response, contradicting the requirement for perfect correlation on average. This concludes the proof.

The precise (i.e. tight) bound is

$$A \leq \frac{5}{6}. \tag{4.8}$$

We provide the full proof in Appendix A.1. This is our noncontextuality inequality.

4.3.2 Quantum violation of the inequality

Quantum theory predicts there is a set of preparations and measurements on a qubit having the supposed properties and achieving $A = 1$, the logical maximum. (These preparations and measurements are precisely those considered in Section 2.3.3). Take the M_t to be represented by the observables $\boldsymbol{\sigma} \cdot \boldsymbol{n}_t$ where $\boldsymbol{\sigma}$ is the vector of Pauli operators and the unit vectors $\{\boldsymbol{n}_1, \boldsymbol{n}_2, \boldsymbol{n}_3\}$ are separated by 120° in the $\boldsymbol{x}$–$\boldsymbol{z}$ plane of the Bloch sphere of qubit states [5]. The $P_{t,b}$ are the eigenstates of these observables, where we associate the positive eigenstate $|+\boldsymbol{n}_t\rangle\langle+\boldsymbol{n}_t|$ with $b = 0$. To see that the statistical equivalence of Equation (4.2) is satisfied, it suffices to note that

$$\frac{1}{3}|+\boldsymbol{n}_1\rangle\langle+\boldsymbol{n}_1| + \frac{1}{3}|+\boldsymbol{n}_2\rangle\langle+\boldsymbol{n}_2| + \frac{1}{3}|+\boldsymbol{n}_3\rangle\langle+\boldsymbol{n}_3| = \frac{1}{2}\mathbb{I}, \tag{4.9}$$

and to recall that for any density operator ρ, $\mathrm{tr}(\rho\frac{1}{2}\mathbb{I}) = \frac{1}{2}$. To see that the statistical equivalence of Equation (4.4) is satisfied, it suffices to note that for all pairs $t, t' \in \{1, 2, 3\}$,

$$\begin{aligned} &\tfrac{1}{2}|+\boldsymbol{n}_t\rangle\langle+\boldsymbol{n}_t| + \tfrac{1}{2}|-\boldsymbol{n}_t\rangle\langle-\boldsymbol{n}_t| \\ &= \tfrac{1}{2}|+\boldsymbol{n}_{t'}\rangle\langle+\boldsymbol{n}_{t'}| + \tfrac{1}{2}|-\boldsymbol{n}_{t'}\rangle\langle-\boldsymbol{n}_{t'}|, \end{aligned} \tag{4.10}$$

which asserts that the average density operator for each value of t is the same, and therefore leads to precisely the same statistics for all measurements. Finally, it is clear that the outcome of the measurement of $\boldsymbol{\sigma} \cdot \boldsymbol{n}_t$ is necessarily perfectly correlated with whether the state was $|+\boldsymbol{n}_t\rangle\langle+\boldsymbol{n}_t|$ or $|-\boldsymbol{n}_t\rangle\langle-\boldsymbol{n}_t|$, so that $A = 1$.

These quantum measurements and preparations are what we seek to implement experimentally, so we refer to them as *ideal*, and denote them by M_t^{i} and $P_{t,b}^{\mathrm{i}}$.

Note that our noncontextuality inequality can accommodate noise in both the measurements and the preparations, up to the point where the average of $p(X = b|M_t, P_{t,b})$ drops below $\frac{5}{6}$. It is in this sense that our inequality does not presume the idealization of noiseless measurements.

4.3.3 Contending with the lack of exact operational equivalence

The actual preparations and measurements in the experiment, which we call the *primary* procedures and denote by $P_{1,0}^{\mathrm{p}}$, $P_{1,1}^{\mathrm{p}}$, $P_{2,0}^{\mathrm{p}}$, $P_{2,1}^{\mathrm{p}}$, $P_{3,0}^{\mathrm{p}}$, $P_{3,1}^{\mathrm{p}}$ and M_1^{p}, M_2^{p}, M_3^{p}, almost certainly deviate from the ideal versions and consequently their mixtures, that is, P_1^{p}, P_2^{p}, P_3^{p} and M_*^{p}, fail to achieve strict equality in Equations (4.2) and (4.4).

We solve this problem as follows. From the outcome probabilities on the six primary preparations, one can infer the outcome probabilities on the entire family of probabilistic mixtures of these. It is possible to find within this family many sets of six preparations, $P_{1,0}^{\mathrm{s}}$, $P_{1,1}^{\mathrm{s}}$, $P_{2,0}^{\mathrm{s}}$, $P_{2,1}^{\mathrm{s}}$, $P_{3,0}^{\mathrm{s}}$, $P_{3,1}^{\mathrm{s}}$, which define mixed preparations P_1^{s}, P_2^{s}, P_3^{s} that satisfy the operational equivalences of Equation (4.4) *exactly*. We call the $P_{t,b}^{\mathrm{s}}$ *secondary* preparations. We can define secondary measurements M_1^{s}, M_2^{s}, M_3^{s} and their uniform mixture M_*^{s} in a similar fashion. The essence of our approach, then, is to identify such secondary sets of procedures and use *these* to calculate A. If quantum theory correctly models our experiment, then we expect to get a value of A close to 1 if and only if we can find suitable secondary procedures that are close to the ideal versions.

To test the hypothesis of noncontextuality, one must allow for the possibility that the experimental procedures *do not* admit of a quantum model. Nonetheless, for pedagogical purposes, we will first provide the details of how one would construct the secondary sets under the assumption that all the experimental procedures do admit of a quantum model.

In Fig. 4.1, we describe the construction of secondary preparations in a simplified example of six density operators that deviate from the ideal states only *within* the $\boldsymbol{x}$–$\boldsymbol{z}$ plane of the Bloch sphere.

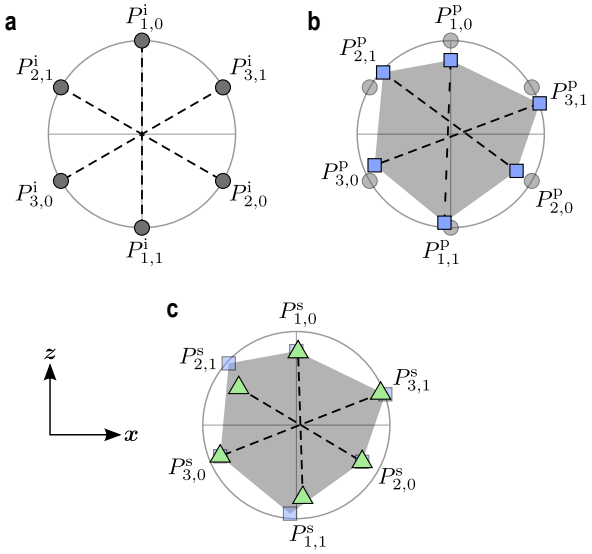

Figure 4.1: Solution to the problem of inexact operational equivalences. Here, we illustrate our solution for the case of preparations under the simplifying assumption that these are confined to the $\boldsymbol{x}$–$\boldsymbol{z}$ plane of the Bloch sphere. For a given pair, $P_{t,0}$ and $P_{t,1}$, the midpoint along the line connecting the corresponding points represents their equal mixture, P_t. **a**, The target preparations $P^{\mathrm{i}}_{t,b}$, with the coincidence of the midpoints of the three lines illustrating that they satisfy the operational equivalence (4.4) exactly. **b**, Illustration of how errors in the experiment (exaggerated in magnitude) will imply that the realized preparations $P^{\mathrm{p}}_{t,b}$ (termed primary) will deviate from the ideal. The lines indicate that not only do these preparations fail to satify the operational equivalence (4.4), but since the lines do not meet, no mixtures of the $P^{\mathrm{p}}_{t,0}$ and $P^{\mathrm{p}}_{t,1}$ can be found at a single point independent of t. The set of preparations corresponding to probabilistic mixtures of the $P^{\mathrm{p}}_{t,b}$ are depicted by the grey region. **c**, Secondary preparations $P^{\mathrm{s}}_{t,b}$ have been chosen from this grey region, with the coincidence of the midpoints of the three lines indicating that the operational equivalence (4.4) has been restored. Note that we require only that the mixtures of the three pairs of preparations be the same, not that they correspond to the completely mixed state.

In practice, the six density operators realized in the experiment will not quite lie in a plane. We use the same idea to contend with this, but with one refinement: we supplement our set of ideal preparations with two additional ones, denoted $P^{\mathrm{i}}_{4,0}$ and $P^{\mathrm{i}}_{4,1}$ corresponding to the two eigenstates of $\boldsymbol{\sigma} \cdot \boldsymbol{y}$. The two procedures that are actually realized in the experiment are denoted $P^{\mathrm{p}}_{4,0}$ and $P^{\mathrm{p}}_{4,1}$ and are considered supplements to the primary set. We then search for our six secondary preparations among the probabilistic mixtures of this supplemented set of primaries rather than among the probabilistic mixtures of the original set. Without this refinement, it can happen that one cannot find six secondary preparations that are close to the ideal versions, as we explain in Appendix A.2.

The scheme for defining secondary measurement procedures is also described in Appendix A.2. Analogously to the case of preparations, one contends with deviations from the plane by supplementing the ideal set with the observable $\boldsymbol{\sigma} \cdot \boldsymbol{y}$.

Note that in order to identify which density operators have been realized in an experiment, the set of measurements must be complete for state tomography [24]. Similarly, to identify which sets of effects have been realized, the set of preparations must be complete for measurement tomography [32]. However, the original ideal sets fail to be tomographically complete because they are restricted to a plane of the Bloch sphere, and an effective way to complete them is to add the observable $\boldsymbol{\sigma} \cdot \boldsymbol{y}$ to the measurements and its eigenstates to the preparations. Therefore, even if we did not already need to supplement these ideal sets for the purpose of providing greater leeway in the construction of the secondary procedures, we would be forced to do so in order to ensure that one can achieve full tomography.

The relevant procedure here is not quite state tomography in the usual sense, since we want to allow for systematic errors in the measurements as well as the preparations. Hence the task [91, 92] is to find a set of qubit density operators, $\rho_{t,b}$, and POVMs, $\{E_{X|t}\}$, that together make the measured data as likely as possible (we cannot expect $\mathrm{tr}(\rho_{t,b}E_{X|t})$ to match the measured relative frequencies exactly due to the finite number of experimental runs).

To analyze our data in a manner that does not prejudice which model—noncontextual, quantum, or otherwise—does justice to it, we must search for representations of the preparations and measurements not amongst density operators and sets of effects, but rather their more abstract counterparts in the formalism of generalised probabilistic theories [50, 51] (GPTs), called generalised states and effects. The assumption that the system is a qubit is replaced by the strictly weaker assumption that three two-outcome measurements are tomographically complete. (In GPTs, a set of measurements are called tomographically complete if their statistics suffice to determine the state.) We take these states and effects as estimates of our primary preparations and measurements, and we

define our estimate of the secondary procedures in terms of these, which in turn are used to calculate our estimate for A. We explain how the raw data is fit to a set of generalised states and effects in Appendix A.3. We characterize the quality of this fit with a χ^2 test.

4.3.4 Experiment

We use the polarization of single photons to test our noncontextuality inequality. The set-up, shown in Figure 4.2, consists of a heralded single-photon source [93, 94, 95], polarization-state preparation and polarization measurement. We generate photons using spontaneous parametric downconversion and prepare eight polarization states using a polarizer followed by a quarter-wave plate (QWP) and half-wave plate (HWP). The four polarization measurements are performed using a HWP, QWP and polarizing beamsplitter. Photons are counted after the beamsplitter and the counts are taken to be fair samples of the true probabilities for obtaining each outcome for every preparation-measurement pair. Since the orientations of the preparation waveplates lead to small deflections of the beam, some information about the preparation gets encoded spatially, and similarly the measurement waveplates create sensitivity to spatial information; coupling the beam into the single-mode fibre connecting the state-preparation and measurement stages of the experiment removes sensitivity to these effects. For a single experimental run we implement each preparation-measurement pair for 4s (approximately 10^5 counts). We performed 100 such runs.

Preparations are represented by vectors of raw data specifying the relative frequencies of outcomes for each measurement, uncertainties on which are calculated assuming Poissonian uncertainty in the photon counts. For each run, the raw data is fit to a set of states and effects in a GPT in which three binary-outcome measurements are tomographically complete. This is done using a total weighted least-squares method [96, 97]. The average χ^2 over the 100 runs is 3.9 ± 0.3, agreeing with the expected value of 4, and indicating that the model fits the data well (see Appendix A.3). The fit returns a 4×8 matrix that serves to define the 8 GPT states and 4 GPT effects, which are our estimates of the primary preparations and measurements. The column of this matrix associated to the t, b preparation, which we denote $\mathbf{P}^{\mathrm{p}}_{t,b}$, specifies our estimate of the probabilities assigned by the primary preparation $P^{\mathrm{p}}_{t,b}$ to outcome '0' of each of the primary measurements. The raw and primary data are compared in Figure 4.3. The probabilities are indistinguishable on this scale. We plot the probabilities for P_1, P_2, and P_3 in Figure 4.4**a** on a much finer scale. We then see that the primary data are within error of the raw data, as expected given the high quality of the fit to the GPT. However, the operational equivalences of

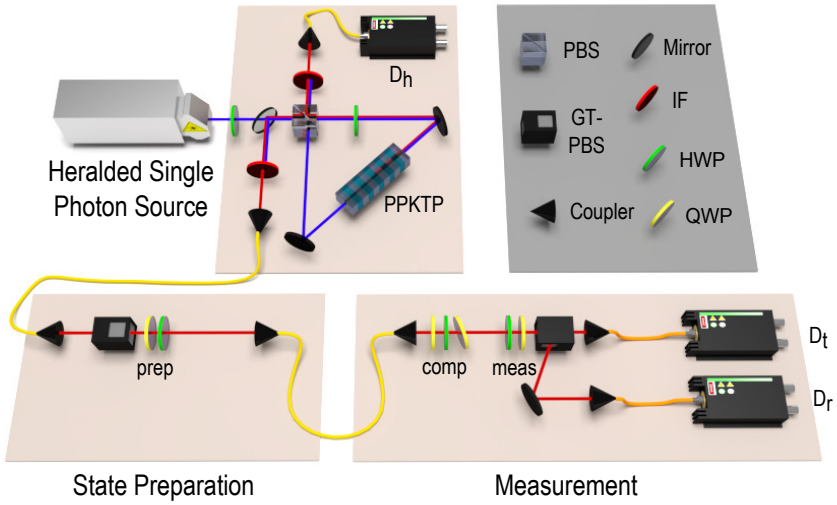

Figure 4.2: The experimental setup. Polarization-separable photon pairs are created via parametric downconversion, and detection of a photon at D_{h} heralds the presence of a single photon. The polarization state of this photon is prepared with a polarizer and two waveplates (prep). A single-mode fibre is a spatial filter that decouples beam deflections caused by the state-preparation and measurement waveplates from the coupling efficiency into the detectors. Three waveplates (comp) are set to undo the polarization rotation caused by the fibre. Two waveplates (meas), a polarizing beamsplitter, and detectors D_{r} and D_{t} perform a two-outcome measurement on the state. PPKTP, periodically-poled potassium titanyl phosphate; PBS, polarizing beamsplitter; GT-PBS, Glan-Taylor polarizing beamsplitter; IF, interference filter; HWP, half-waveplate; QWP, quarter-waveplate.

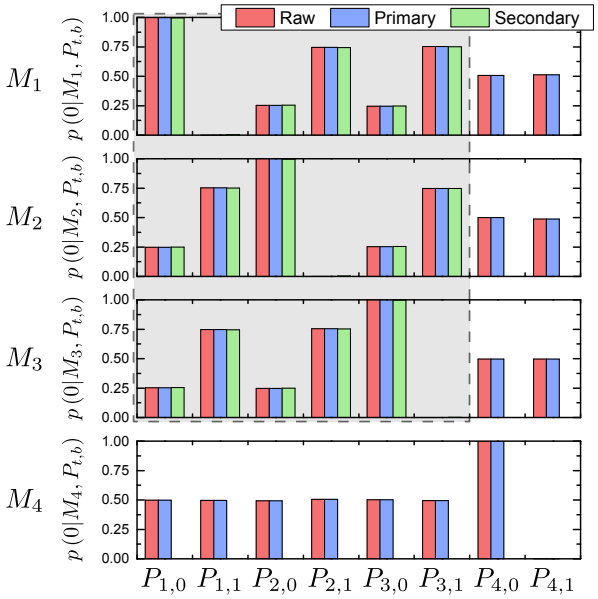

Figure 4.3: Outcome probabilities for each measurement-preparation pair. For each such pair, we report the probability of obtaining outcome 0 in the measurement. Red bars are relative frequencies calculated from the raw counts, blue bars are our estimates of the outcome probabilities of the primary measurements on the primary preparations obtained from a best-fit of the raw data, and green bars are our estimates of the outcome probabilities of the secondary measurements on the secondary preparations. The shaded grey background highlights the measurements and preparations for which secondary procedures were found. Error bars are not visible on this scale, neither are discrepancies between the obtained probabilities and the ideal values thereof, which are at most 0.013; statistical error due to Poissonian count statistics is at most 0.002.

Equations (4.2) and (4.4) are not satisfied by our estimates of the primary preparations and measurements, illustrating the need for secondary procedures.

We define the six secondary preparations as probabilistic mixtures of the eight primaries: $\mathbf{P}^{\mathrm{s}}_{t,b} = \sum_{t'=1}^{4}\sum_{b'=0}^{1} u^{t,b}_{t',b'}\mathbf{P}^{\mathrm{p}}_{t',b'}$, where the $u^{t,b}_{t',b'}$ are the weights in the mixture. We maximize $C_{\mathrm{P}} = \frac{1}{6}\sum_{t=1}^{3}\sum_{b=0}^{1} u^{t,b}_{t,b}$ over valid $u^{t,b}_{t',b'}$ subject to the constraint of Equation (4.4), that is, $\frac{1}{2}\sum_b \mathbf{P}^{\mathrm{s}}_{1,b} = \frac{1}{2}\sum_b \mathbf{P}^{\mathrm{s}}_{2,b} = \frac{1}{2}\sum_b \mathbf{P}^{\mathrm{s}}_{3,b}$ (a linear program). A high value of C_{P} ensures each of the six secondary preparations is close to its corresponding primary. Averaging over 100 runs, we find $C_{\mathrm{P}} = 0.9969 \pm 0.0001$, close to the maximum of 1. An analogous linear program to select secondary measurements yields similar results. In Appendix A.2, Tables A.2 and A.3 display the weights that define each secondary preparation and measurement, averaged over 100 runs. Figure 4.3 also displays the outcome probabilities for the secondary procedures, confirming that they are close to ideal. Figure 4.4 demonstrates how our construction enforces the operational equivalences.

We analyzed each experimental run separately and found the degree of correlation $p(X{=}b|M^{\mathrm{s}}_t, P^{\mathrm{s}}_{t,b})$ for each value of t and b. The averages over the 100 runs are shown in Figure 4.5**a** and are all in excess of 0.995. Averaging over t and b yields an experimental value $A = 0.99709 \pm 0.00007$, which violates the noncontextual bound of $5/6 \approx 0.833$ by 2300σ (Figure 4.5**b**).

4.4 Discussion

Using the techniques described here, it is possible to convert proofs of the failure of noncontextuality in quantum theory into experimental tests of noncontextuality that are robust to noise and experimental imprecisions [98, 99]. For any phenomenon, therefore, one can determine which of its operational features are genuinely nonclassical. This is likely to have applications for scientific fields wherein quantum effects are important and for developing novel quantum technologies. The definition of operational equivalence of preparations (measurements) required them to be statistically equivalent relative to a tomographically complete set of measurements (preparations). There are two examples of how the assumption of tomographic completeness is expected *not* to hold exactly in our experiment, even if one grants the correctness of quantum theory. First, our source produces a small multi-photon component. We measure the heralded $g^{(2)}(0)$ of our source [11] to be 0.00105 ± 0.00001 and from this[1] we estimate the ratio of heralded detection events caused by multiple photons to those caused by single photons to be 1:4000. Regardless

[1] In Ref. [2] the heralded $g^{(2)}(0)$ was mistakenly reported as 0.0105 ± 0.0001.

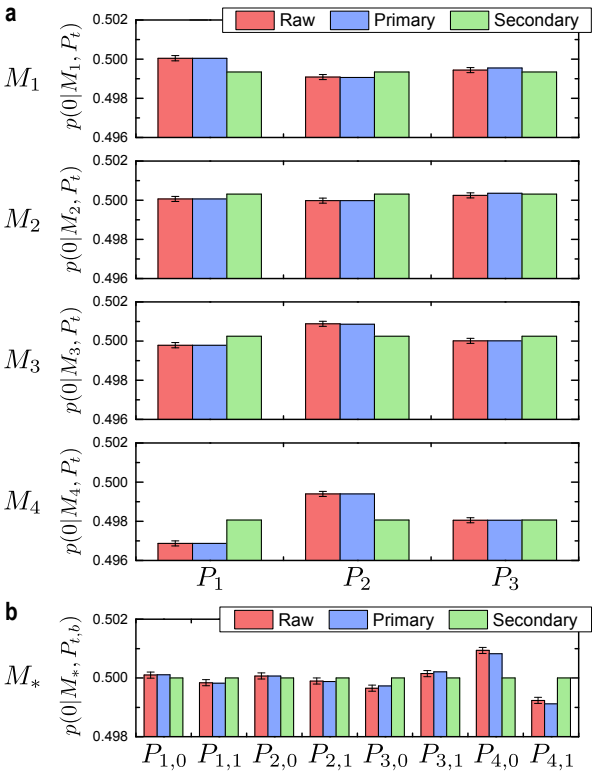

Figure 4.4: Verifying the requisite operational equivalences. Operational statistics for raw, primary, and secondary preparations and measurements, averaged over 100 experimental runs. **a**, The probabilities of the primary measurements (blue bars) differ depending on which of the three mixed preparations P_1^{p}, P_2^{p}, and P_3^{p} are measured. These probabilities are within error of the raw data (red bars), indicating that a generalised probabilistic theory in which three two-outcome measurements are tomographically complete fits the data well. Probabilities for primary measurements on the secondary preparations (green bars) are independent of the preparation, hence the secondary preparations satisfy Equation (4.4). Note that one expects these probabilities to deviate from 0.5. In the example of Figure 4.1**c**, this corresponds to the fact that the intersection of the lines is not the completely mixed state. **b**, Outcome probabilities of measurement M_* on the eight preparations. Red bars are raw data, blue bars are the measurement M_*^{p} on the primary preparations, and green bars are M_*^{s} on the primary preparations. Regardless of the input state, M_*^{s} returns outcome 0 with probability 0.5, hence it is operationally indistinguishable from a fair-coin flip (Equation (4.2)). Error bars in all plots are calculated assuming Poissonian count statistics.

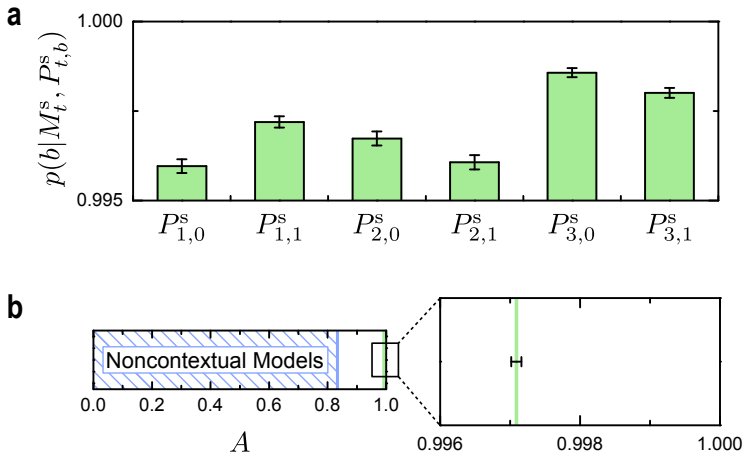

Figure 4.5: Violation of the noncontextuality inequality. **a**, Values of the six degrees of correlation in Equation (4.8), averaged over 100 experimental runs. **b**, Average measured value for A contrasted with the noncontextual bound $A = 5/6$. We find $A = 0.99709 \pm 0.00007$, which violates the noncontextual bound by 2300σ. Error bars in both plots represent the standard deviation in the average of the measured values over the 100 experimental runs.

of the value of A that one presumes for multi-photon events, one can infer that the value of A we would have achieved had the source been purely single-photon could have been less than the value given above by at most 10^{-6}, a difference that does not affect our conclusions. We also expect the assumption to not hold exactly because of the inevitable coupling of the polarization into the spatial degree of freedom of the photon, which could be caused, for example, by a wedge in a waveplate. Indeed, we found that if the spatial filter was omitted from the experiment, our fitting routine returned large χ^2 values, which we attributed to the fact that different angles of the waveplates led to different deflections of the beam. A more abstract worry is that *nature* might conflict with the assumption (and prediction of quantum theory) that three independent binary-outcome measurements are tomographically complete for the polarization of a photon. Our experiment has provided evidence in favour of the assumption insofar as we have fit data from *four* measurements to a theory where three are tomographically complete and found a good χ^2 value for the fit. One can imagine accumulating much more evidence of this sort, but it is difficult to see how any experiment could conclusively vindicate the assumption, given that one can never test all possible measurements. This, therefore, represents the most significant loophole in experimental tests of noncontextuality, and new ideas for how one might seal it or circumvent it represent the new frontier for improving such tests.

4.5 Methods

4.5.1 Preparation procedure

A 20-mW diode laser with a wavelength of 404.7 nm produces photon pairs, one horizontally polarized and the other vertically polarized, via spontaneous parametric down-conversion in a 20-mm type-II PPKTP crystal. The downconversion crystal is inside a Sagnac loop and the pump laser is polarized vertically to ensure it only travels counter-clockwise around the loop. Photon pairs are separated at a polarizing beamsplitter and coupled into two single-mode fibres (SMFs). Vertically-polarized photons are detected immediately at detector D_{h}, heralding the presence of the horizontally-polarized signal photons which emerge from the SMF and pass through a state-preparation stage before they are measured. Herald photons were detected at a rate of 400 kHz. Signal photons emerge from the fibre and pass through a Glan-Taylor polarizing beamsplitter (GT-PBS) which transmits vertically-polarized light. Polarization controllers in the fibre maximize the number of photons which pass through the beamsplitter. A quarter- and half-waveplate set the polarization of the signal photons to one of eight states.

4.5.2 Spatial mode filter

An SMF acts as a spatial mode filter. This filter ensures that information about the angles of the state-preparation waveplates cannot be encoded in the spatial mode of the photons, and that our measurement procedures do not have a response that depends on the spatial mode, but only on polarization as intended. The SMF induces a fixed polarization rotation, so a set of three compensation waveplates are included after the SMF to undo this rotation. It follows that the preparation-measurement pairs implemented in our experiment are in fact a rotated version of the ideal preparation and a similarly-rotated version of the ideal measurement. Such a fixed rotation, however, does not impact any of our analysis.

4.5.3 Measurement procedure

Measurements are performed in four bases, set by a half- and quarter-waveplate. A second GT-PBS splits the light, and both output ports are detected. Due to differences in the coupling and detection efficiencies in each path after the beamsplitter, each measurement consists of two parts. First, the waveplates are aligned such that states corresponding to outcome '0' are transmitted by the GT-PBS, and the number of heralded photons detected

in a two-second window is recorded for each port. Second, the waveplate angles are changed in such a way as to invert the outcomes, so the detector in the reflected port corresponds to outcome '0' and heralded photons are detected for another two seconds. The counts are added together and the probability for outcome '0' is calculated by dividing the number of detections corresponding to outcome '0' by the total number of detection events in the four-second window. The single-photon detection rate at detectors D_{r} and D_{t} depends on the measurement settings. In the transmissive and reflective ports of the measurement GT-PBS photons were detected at maximum rates of 330 kHz and 250 kHz, respectively. Coincident detection events between herald photons and the transmissive and reflective ports of the measurement GT-PBS were up to 22 kHz and 16 kHz, respectively.

Chapter 5

Experimentally bounding deviations from quantum theory within the framework of generalised probabilistic theories

Notes and acknowledgements

This chapter contains material from work that has appeared on the arXiv as [1]:

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," *arXiv:1710.05948*, (2017).

Author contributions

Robert Spekkens conceived of the experiment.

Michael Mazurek and **Kevin Resch** designed the experiment.

Michael Mazurek performed the experiment and collected the data.

All authors discussed how to interpret the data and contributed to the final draft of the manuscript.

Michael Mazurek and **Matthew Pusey** implemented the numerical data analysis.
Michael Mazurek and **Robert Spekkens** wrote the initial draft of the manuscript.

5.1 Overview

Many experiments in the field of quantum foundations seek to adjudicate between quantum theory and speculative alternatives to it. To do so, one must analyse the experimental data in a manner that does not presume the correctness of the quantum formalism. The mathematical framework of generalized probabilistic theories (GPTs) provides a means of doing so. We present a scheme for determining what GPTs are consistent with a given set of experimental data. It proceeds by performing tomography on the preparations and measurements in a self-consistent manner, i.e., without presuming a prior characterization of either. We illustrate the scheme by analyzing experimental data for a large set of preparations and measurements on the polarization degree of freedom of a single photon. We find that the smallest and largest GPT state spaces consistent with our data are a pair of polytopes, each approximating the shape of the Bloch Sphere and having a volume ratio of 0.977 ± 0.001, which provides a quantitative bound on the scope for deviations from quantum theory. We also demonstrate how our scheme can be used to bound the extent to which nature might be more nonlocal than quantum theory predicts, as well as the extent to which it might be more or less contextual. Specifically, we find that the maximal violation of the CHSH inequality can be at most $(1.3 \pm 0.1)\%$ greater than the quantum prediction, and the maximal violation of a particular noncontextuality inequality can not differ from the quantum prediction by more than this factor on either side.

5.2 Introduction

Despite the empirical successes of quantum theory, it may one day be supplanted by a novel, post-quantum theory.[1] Many researchers have sought to anticipate what such a theory might look like based on theoretical considerations, in particular, by exploring how various natural physical principles narrow down the scope of possibilities in the landscape of all physical theories (see [100] and references therein). In this article, we consider a complementary problem: how to narrow down the scope of possibilities directly from experimental data.

[1] Indeed, the fact that it has not yet been unified with general relativity is often cited as evidence for the need to go beyond quantum theory.

Most experiments in the field of quantum foundations aim to adjudicate between quantum theory and some speculative alternative to it. They seek to constrain (and perhaps uncover) deviations from the quantum predictions. Although a few proposed alternatives to quantum theory *can* be articulated within the quantum formalism itself, such as models which posit intrinsic decoherence [101, 102, 103, 104], most are more radical. Examples include Almost Quantum Theory [105, 106], theories with higher-order interference (or of higher-order in the sense of Ref. [50])[107, 108, 109, 110, 111, 112], and modifications to quantum theory involving the quaternions [113, 114, 115, 116].

In order to assess whether experimental data provides any evidence for a given proposal (and against quantum theory), it is clearly critical that one *not* presume the correctness of quantum theory in the analysis. Therefore it is inappropriate to use the quantum formalism to model the experiment. A more general formalism is required. Furthermore, it would be useful if rather than implementing dedicated experiments for each proposed alternative to quantum theory, one had a technique for directly determining the experimentally viable regions in the landscape of all possible physical theories. The framework of generalized probabilistic theories (GPTs) provides the means to meet both of these challenges.

This framework adopts an operational approach to describing the content of a physical theory. It has been developed over the past fifteen years in the field of quantum foundations (see [50, 51, 52, 53] in particular, as well as [117, 54, 118, 119, 56, 120, 121, 122, 106]), continuing a long tradition of such approaches [123, 124, 125, 126]. It is *operational* because it takes the content of a physical theory to be merely what it predicts for the probabilities of outcomes of measurements in an experiment.

The GPT framework makes only very weak assumptions (which are arguably unavoidable if an operationalist's conception of an experiment is to be meaningful). One is that experiments have a modular form, such that one part of an experiment can be varied independently of another, such as preparations and measurements for instance; another is that it is possible to repeat a given experimental configuration in such a way that it constitutes an i.i.d. source of statistical data. Beyond this, however, it is completely general. It has been used extensively to provide a common language for describing and comparing abstract quantum theory, classical probability theory, and many foils to these, such as quantum theory over the real or quaternionic fields[116], theories with higher-order interference[127, 128, 129], Boxworld [51, 56], or Almost Quantum Theory [106].

Using this framework, we propose a technique for analyzing experimental data that allows researchers to overcome their implicit quantum bias — the tendency of viewing all experiments through the lens of quantum concepts and formalism — and take a theory-neutral perspective on the data.

Despite the fact that the GPT formalism is ideally suited to the task, to our knowledge, it has not previously been applied to the analysis of experimental data, (with the exception of Ref. [2], which applied it to an experimental test of universal noncontextuality and which inspired the present work).

In this paper we answer the question: given a set of experimental data, how does one find the set of GPTs that could have generated the data? We call this the "GPT inference problem". Solving the problem requires implementing the GPT analogue of quantum tomography. Quantum tomography experiments that have sought to characterize unknown states have typically presumed that the measurements are already well-characterized [130, 131, 132, 133, 24, 134, 135], and those have sought to characterize unknown measurements have presumed that the states are known [30, 32]. If one has no prior knowledge of either the states or the measurements, then one requires a tomography scheme that can characterize them both based on their interplay. We call such a tomographic scheme *self-consistent*. To solve the GPT inference problem, we introduce such a self-consistent tomography scheme within the framework of GPTs.

We illustrate the use of our technique with an experiment on the polarization degree of freedom of a single photon. For each of a large number of preparations, we perform a large number of measurements, and from the data we use our tomography scheme to infer a GPT characterization of both the preparations and the measurements. From this characterization, we place bounds (at the 1% level) on how much the true theory describing our experiment might deviate from quantum theory in various respects. In addition, we draw explicit quantitative conclusions about three types of deviations from quantum theory.

A popular axiom in reconstructions of quantum theory, termed the *no-restriction hypothesis* [52] asserts that if some measurement is logically possible (i.e., it gives positive probabilities for all states in the theory) then it should be physically realizable. A failure of the no restriction hypothesis, therefore, constitutes a departure from quantum theory. We put quantitative bounds on the degree of this failure, that is, on the potential gap between the set of measurements that are physically realizable and those that are logically possible.

We also put an upper bound on the amount by which nature might violate Bell inequalities in excess of the amount predicted by quantum theory. Specifically, for the CHSH inequality [48], we show that for photon polarization any greater-than-quantum degree of violation is no more than 1.3% higher than the quantum bound. To our knowledge, this is the first proposal for how to obtain an experimental *upper* bound on the degree of Bell inequality violation in nature.

In a similar vein, we consider noncontextuality inequalities. These are akin to Bell

inequalities, but test the hypothesis of universal noncontextuality [38] rather than local causality. Here, our technique provides both an upper and a lower bound on the degree of violation. For a particular noncontextuality inequality, described in Ref. [61], we find that the true value of the violation cannot be greater than or less than the quantum value by more than 1.3%.

5.3 The GPT inference problem

As we saw in Chapter 3, the true GPT state and effect spaces for a physical system, $\mathcal{S}$ and $\mathcal{E}$, are theoretical abstractions, describing the full set of GPT state and effect vectors that could be in principle be realized experimentally if one could eliminate all noise. However, the ideal of noiselessness is never achieved. Therefore, the GPT state and effect vectors describing the preparation and measurement effects realized in any experiment are necessarily bounded away from the extremal elements of $\mathcal{S}$ and $\mathcal{E}$. Geometrically, the realized GPT state and effect spaces will be contracted relative to their true counterparts.

There is another way in which the experiment necessarily differs from the theoretical abstraction: it may be impossible for the set of experimental configurations in a real experiment to probe *all* possible experimental configurations allowed by the GPT. For instance, for quantum theory there are an *infinite* number of convexly extremal preparations and measurements even for a single qubit, while a real experiment can only implement a finite number of each.

Because we assume convex closure, the realized GPT state and effect spaces will be polytopes. If the experiment probes a sufficiently dense sample of the preparations and measurements allowed by the GPT, then the shapes of these polytopes ought to resemble the shapes of their true counterparts.

We term the convex hull of the GPT states that are actually realized in an experiment the *realized GPT state space*, and denote it by $\mathcal{S}_{\text{realized}}$. Because every preparation is noisier than the ideal version thereof, this will necessarily be *strictly* contained within the true GPT state space $\mathcal{S}$. Similarly, we term the diamond defined by the GPT measurement effects that are actually realized in an experiment the *realized GPT effect space*, and denote it $\mathcal{E}_{\text{realized}}$. Again, we expect it to be strictly contained within $\mathcal{E}$. By dualization, $\mathcal{S}_{\text{realized}}$ defines the set of GPT effect vectors that are logically consistent with the realized preparations, which we denote by $\mathcal{E}_{\text{consistent}}$, that is, $\mathcal{E}_{\text{consistent}} \equiv \text{dual}(\mathcal{S}_{\text{realized}})$. Similarly, the set of GPT state vectors that are logically consistent with the realized measurement effects is $\mathcal{S}_{\text{consistent}} \equiv \text{dual}(\mathcal{E}_{\text{realized}})$.

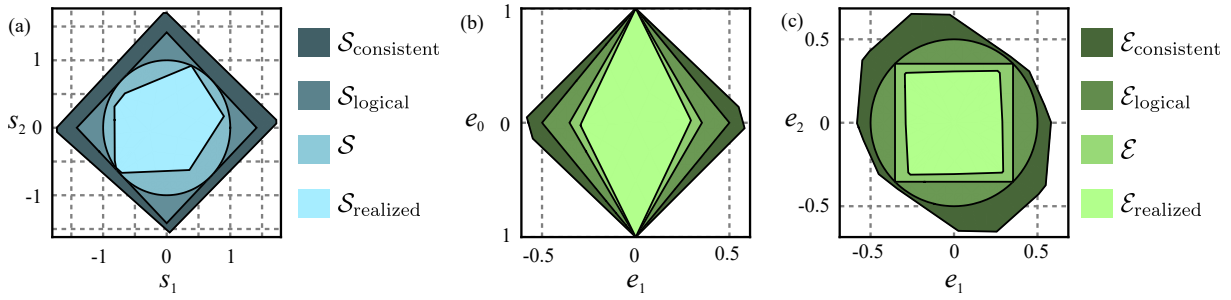

Figure 5.1: An illustration of the inclusion relations among the different spaces of states and effects considered in this chapter. We use a generic $k = 3$ example for ease of depicting set inclusions. (a) The different spaces of states. (b),(c) The 2-d projections of the different spaces of effects. The GPT specifies a space of true states, $\mathcal{S}$, and effects, $\mathcal{E}$. From these, one can find the sets of logically possible states, $\mathcal{S}_{\text{logical}}$, and effects $\mathcal{E}_{\text{logical}}$. $\mathcal{E}_{\text{logical}}$ is the dual of $\mathcal{S}$, and it represents all effects which return probabilities between 0 and 1 when applied to every possible state in $\mathcal{S}$. Similarly, $\mathcal{S}_{\text{logical}}$ is the dual of $\mathcal{E}$. The logical state (effect) space must always contain the true state (effect) space. The spaces $\mathcal{S}_{\text{realized}}$ and $\mathcal{E}_{\text{realized}}$ are the GPT representations of the preparations and measurement effects actually realized in the experiment. As any real experiment necessarily contains a finite amount of noise, $\mathcal{S}_{\text{realized}}$ will always be contained within $\mathcal{S}$, and $\mathcal{E}_{\text{realized}}$ will always be contained within $\mathcal{E}$. $\mathcal{E}_{\text{consistent}}$ is the dual of $\mathcal{S}_{\text{realized}}$ (and thus will always contain $\mathcal{E}_{\text{logical}}$), and it represents all effects that are logically consistent with the set of states realized in the experiment. Similarly, $\mathcal{S}_{\text{consistent}}$ will always contain $\mathcal{S}_{\text{logical}}$ as it is the dual of $\mathcal{E}_{\text{realized}}$.

Suppose one has knowledge of the realized GPT state and effect spaces $\mathcal{S}_{\text{realized}}$ and $\mathcal{E}_{\text{realized}}$ for some experiment. What can one then infer about $\mathcal{S}$ and $\mathcal{E}$? The answer is that $\mathcal{S}$ can be any convex set of GPT states that lies strictly between $\mathcal{S}_{\text{realized}}$ and $\mathcal{S}_{\text{consistent}}$. For every such possibility for $\mathcal{S}$, $\mathcal{E}$ could be any diamond of GPT effects that lies between $\mathcal{E}_{\text{realized}}$ and $\text{dual}(\mathcal{S}) \subset \mathcal{E}_{\text{consistent}}$. These inclusion relations are depicted in Fig. 5.1.

The larger the gap between $\mathcal{S}_{\text{realized}}$ and $\mathcal{S}_{\text{consistent}}$, the more choices of $\mathcal{S}$ and $\mathcal{E}$ there are that are consistent with the experimental data. An example helps illustrate the point. Suppose that one found $\mathcal{S}_{\text{realized}}$ and $\mathcal{E}_{\text{realized}}$ to be the GPT state and effect spaces depicted in Fig. 3.1(d). In this case $\mathcal{S}_{\text{realized}}$ is represented by the blue octahedron in Fig. 3.1(d)(i), and $\mathcal{E}_{\text{realized}}$ is the green diamond with an octahedral base depicted in Fig. 3.1(d)(ii-iii). The wireframe cube in Fig. 3.1(d)(i) is the space of states $\mathcal{S}_{\text{consistent}}$ that is the dual of $\mathcal{E}_{\text{realized}}$, and the wireframe diamond with a cubic base in Fig. 3.1(d)(ii-iii) is the space of effects $\mathcal{E}_{\text{consistent}}$ that is the dual of $\text{S}_{\text{realized}}$. Which GPTs are candidates for the true GPT in this case? The answer is: those whose state space contains the blue octahedron

and is contained by the wireframe cube in Fig. 3.1(d)(i) and whose effect space contains the green diamond with the octohedral base in Fig. 3.1(d)(ii)-(iii) (the consistency of the effect space with the state space is a given if one grants that the pair is a valid GPT). By visual inspection of Fig. 3.1(a) and Fig. 3.1(c), it is clear that the GPTs representing both quantum theory and Boxworld are consistent with this data. The GPT for a classical 4-level system (i.e. the $k = 4$ generalization of the classical bit in Fig. 3.1(b) [121]) is as well.

When there is a large gap between $\mathcal{S}_{\text{realized}}$ and $\mathcal{S}_{\text{consistent}}$, it is important to consider the possibility that this is due to a shortcoming in the experiment and that probing more experimental configurations will reduce it. For instance, if an experiment on a 2-level system was governed by quantum theory, but the experimenter only considered experimental configurations involving eigenstates of Pauli operators, then $\mathcal{S}_{\text{realized}}$ and $\mathcal{E}_{\text{realized}}$ would be precisely those of the example we have just described, implying many possibilities besides quantum theory for the true GPT. However, further experimentation would reveal that this seemingly large scope for deviations from quantum theory was merely an artifact of probing a too-sparse set of configurations. Only if one continually fails to close the gap between $\mathcal{S}_{\text{realized}}$ and $\mathcal{S}_{\text{consistent}}$, in spite of probing the greatest possible variety of experimental configurations, should one consider the possibility that in fact $\mathcal{S} \simeq \mathcal{S}_{\text{realized}}$ and $\mathcal{E} \simeq \mathcal{E}_{\text{realized}}$ and that the true GPT fails to satisfy the no-restriction hypothesis. By contrast, if the gap between $\mathcal{S}_{\text{realized}}$ and $\mathcal{S}_{\text{consistent}}$ is very small, the experiment has found a tightly constrained range of possibilities for the true GPT, and it successfully rules out a large class of alternative theories.

5.4 Self-consistent tomography in the GPT framework

We have just seen that any real experiment defines a set of realized GPT states, $\mathcal{S}_{\text{realized}}$, and a set of realized GPT effects, $\mathcal{E}_{\text{realized}}$, and it is from these that one can infer the scope of possibilities for the true spaces, $\mathcal{S}$ and $\mathcal{E}$, and thus the scope of possibilities for deviations from quantum theory.

But how can one estimate $\mathcal{S}_{\text{realized}}$ and $\mathcal{E}_{\text{realized}}$ from experimental data? In other words, how can one implement tomography within the GPT framework? This is the problem whose solution we now describe. The steps in our scheme are outlined in Fig. 5.2.

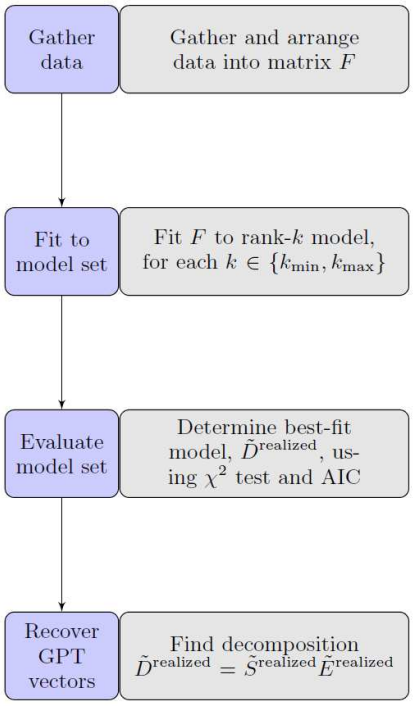

Figure 5.2: Overview of the self-consistent GPT tomography procedure. We begin with the experimental data, finite-run relative frequencies for each configuration realized in the experiment, and arrange it into a matrix, F, which is a noisy version of the matrix of true probabilities, D^{realized}. To estimate the dimension, k, of the data, we find the rank-k matrix which best fits F for a set of values of k. We call this set of best-fit rank-k matrices the *candidate model set*. A statistical analysis on the candidate model set (using the χ^2 goodness-of-fit test and the Akaike information criterion) determines the value of k that gives us the best fit, and therefore which of the candidate models is the best approximation to D^{realized}. We denote this best approximation by $\tilde{D}^{\text{realized}}$. We find a decomposition $\tilde{D}^{\text{realized}} = \tilde{S}^{\text{realized}}\tilde{E}^{\text{realized}}$, in order to estimate the spaces of states and effects realized in the experiment. Each row of $\tilde{S}^{\text{realized}}$ is a GPT state vector representing one of the preparation procedures in the experiment, and each column of $\tilde{E}^{\text{realized}}$ is a GPT effect vector representing one of the measurement procedures. This completes the GPT tomography procedure.

5.4.1 Accumulating evidence for tomographic completeness

Our scheme only works if the data acquired contains sufficient information to fully characterize any preparation or measurement in the GPT. It is imperative, therefore, to ensure that the sets of preparations and measurements in one's experiment are tomographically complete. Because one cannot presume the correctness of quantum theory, however, one does not have any theoretical grounds for deciding which sets of measurements (preparations) are tomographically complete for a given system. If one were to implement *all* extremal measurements and *all* extremal preparations on the system, this would clearly be sufficient, but in cases where there is a continuum of extremal elements (as in quantum theory), this is impossible. Therefore, the best one can do is to implement a set of measurements and preparations on the system that is as large and as diverse as possible. By doing so, one can certainly build up evidence in favour of tomographic completeness. Every novel preparation (measurement) whose statistics are well predicted by those in the putative tomographically complete set adds to the evidence. Nonetheless, one can never rule out the possibility that tomorrow a novel type of preparation (measurement) procedure will be identified whose statistics are *not* predicted by those in the putative tomographically complete set, thereby demonstrating that the set was not tomographically complete after all. As such, any conclusion of tomographic completeness is always tentative.

However, as Popper emphasized, *all* scientific claims are vulnerable to being falsified and therefore have a tentative status [136]. We are therefore recommending to treat the hypothesis that a given set of measurements and a given set of preparations are tomographically complete as Popper recommends treating any scientific hypothesis: one should try one's best to falsify it and as long as one fails to do so, the hypothesis stands.

Building evidence in favour of tomographic completeness is a critical step in our scheme because the validity of all of the conclusions rests upon it.

5.4.2 Inferring best-fit probabilities from finite-run statistics

We suppose that, for a given system, the experimenter makes use of a finite number m of preparation procedures (P_i, $i \in \{1, \cdots, m\}$) and a finite number, n, of binary-outcome measurement procedures (M_j, $j \in \{1, \cdots, n\}$). We denote the outcome of each measurement by $a \in \{0, 1\}$. For each choice of preparation and measurement, (P_i, M_j), the experimenter records the outcome of the measurement in a large number of runs and computes the relative frequency with which a given outcome a occurs, denoted $f(a|P_i, M_j)$. For the binary-outcome measurements under consideration, it is sufficient to specify $f(0|P_i, M_j)$ for each pair (P_i, M_j), because $f(1|P_i, M_j)$ is then fixed by normalization.

The set of all experimental data, therefore, can be encoded in an $m \times n$ matrix F, whose (i, j)th component is $f(0|P_i, M_j)$.

The relative frequency $f(0|P_i, M_j)$ one measures will not coincide exactly with the probability $p(0|P_i, M_j)$ from which it is assumed that the outcome in each run is sampled.[2] Rather, $f(0|P_i, M_j)$ is merely a noisy approximation to $p(0|P_i, M_j)$. The statistical variation in $f(0|P_i, M_j)$ can, however, be estimated from the experiment.

It follows that the matrix F extracted from the experimental data is merely a noisy approximation to the matrix D^{realized} that encodes the predictions of the GPT for the mn experimental configurations of interest. Because of the noise, F will generically be full rank, regardless of the rank of D^{realized} [137]. Therefore, the experimentalist is tasked with estimating the $m \times n$ probability matrix D^{realized} given the $m \times n$ data matrix F, where the rank of D^{realized} is a parameter in the fit.

We aim to describe our technique in a general manner, so that it can be applied to any experiment. However, in order to provide a concrete example of its use, we will intersperse our presentation of the technique with details about how it is applied to the particular experiment we have conducted. We begin, therefore, by providing the details of the latter.

5.4.3 Description of the experiment

To illustrate the GPT tomography scheme, we perform an experiment on the polarization degree of freedom of single photons. Pairs of photons are created via spontaneous parametric down-conversion, and the detection of one of these photons, called the herald, indicates the presence of the other, called the signal. We manipulate the polarization of the signal photons with a quarter- and half- waveplate before they are coupled into a single-mode fibre; each preparation is labelled by the angles of these two waveplates.

Upon emerging from the fibre, photons encounter the measurement stage of the experiment, which consists of a quarter- and half-waveplate followed by a polarizing beam splitter with single-photon detectors at each of its output ports. Each measurement is labelled by the angles of the two waveplates preceding the beam splitter.

The frequency of the $a = 0$ outcome is defined as the ratio of the number of signal photon detections in the $a = 0$ output port to the total number of heralded detections.

[2] Note that it is presumed that the outcome variables for the different runs (on a given choice of preparation and measurement) are identically and independently distributed. This assumption could fail, for instance, due to a drift in the nature of the preparation or measurement over the timescale on which the different runs take place, or due to a memory effect that makes the outcomes in different runs correlated. In such cases, one would require a more sophisticated analysis than the one described here.

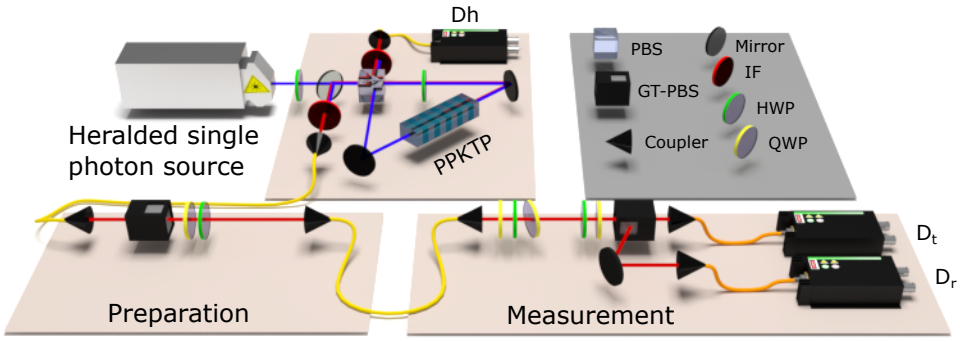

Figure 5.3: The experimental setup. Pairs of polarization-separable single photons are created via spontaneous parametric down-conversion. The herald photon is sent to a detector. The signal photon's polarization travels through a polarizer then a quarter and half waveplate, which prepares its polarization state. The photon is then coupled into single-mode fibre which removes any information which may be encoded in the photon's spatial degree-of-freedom. Three static waveplates undo the polarization rotation caused by the fibre. Two waveplates and a polarizing beamsplitter with detectors in each output port perform a measurement on the photon. One output port is labelled '0', and the other is labelled '1'. Coincident detections between the herald detector, D_h, and the detector in the transmitted port, D_t, are counted, as well as coincidences between D_h and the reflected-port detector D_r. PPKTP: Periodically-poled potassium titanyl phosphate; PBS: Polarizing beamsplitter; GT-PBS: Glan-Thompson polarizing beamsplitter; IF: Interference filter; HWP: Half waveplate; QWP: Quarter waveplate.

We ignore experimental trials in which either the herald or the signal photon is lost by post-selecting on coincident detections, so that our measurements are only performed on normalized states.

We choose $m = 100$ waveplate settings for the preparations, and $n = 100$ waveplate settings for the settings, corresponding to $mn = 10^4$ experimental configurations in all, one for each pairing.

We choose $m = n$ so that the GPT state space and the GPT effect space are equally well characterized. We detect coincidences at a rate of ~ 2250 counts/second, and count coincidences for each preparation-measurement pair for a total of eight seconds, allowing us to achieve a standard deviation on each data point below the 1% level. Because of the additional time it takes to mechanically rotate the preparation and measurement waveplates, it takes approximately 84 hours to acquire data for 10^4 preparation-measurement pairs.

Our method of selecting *which* 100 waveplate settings to use is described in Appendix B.2. Note that although the choice of these settings is motivated by our knowledge of the quantum formalism, our tomographic scheme does not assume the correctness of quantum theory: our reconstruction scheme could have been applied equally well if the waveplate settings had been chosen at random.[3]

The raw frequencies are arranged into the data matrix F. Entry F_{ij} is the frequency at which outcome $a = 0$ was obtained when setting M_j was used to measure a photon prepared with setting P_i. As noted in Sec. 3.2, we adopt a convention wherein M_1 is the unit measurement, implying that the first column of F is a column of 1s. The data matrix for our experiment is presented in Fig. 5.4. As expected, we find that F is full rank.

5.4.4 Estimating the probability matrix D^{realized}

We turn now to the problem of estimating from F the $m \times n$ probability matrix D^{realized}. The first item of business is to estimate the rank of D^{realized}, which is equivalent to estimating the cardinality of the tomographically complete set of preparations (or measurements) of the GPT model of the experiment.

The best estimate of the rank-k probability matrix D^{realized} is the rank-k matrix $\tilde{D}^{\text{realized}}$ that best fits the data matrix F. In other words, $\tilde{D}^{\text{realized}}$ is the rank-k matrix that

[3] An interesting question for future research is how the quality of the GPT reconstruction varies with the particular set of waveplate settings that are considered. In particular, one can ask about the quality of the evidence for quantum theory in the situation wherein the waveplate settings correspond to sampling highly *nonuniformly* over the points on the Bloch sphere.

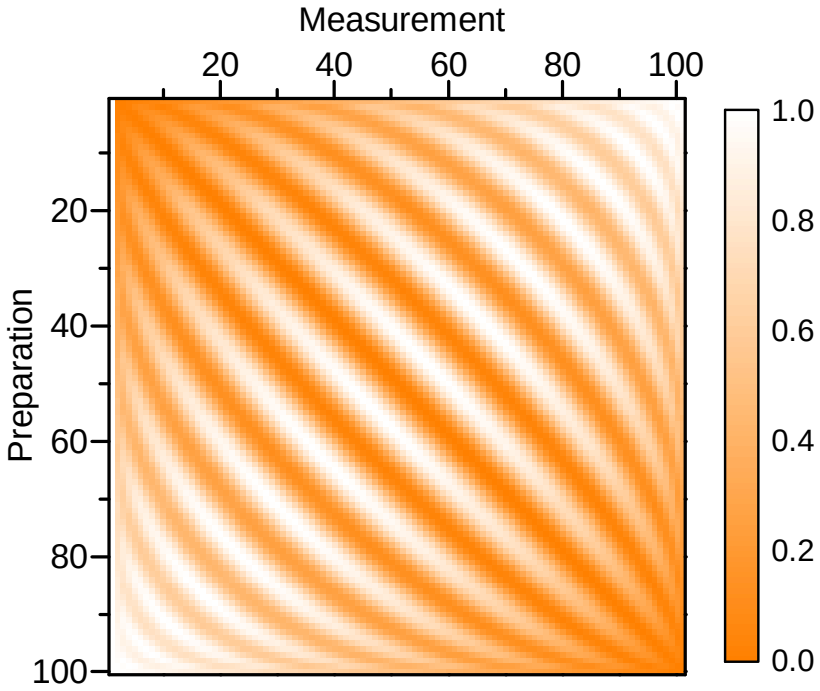

Figure 5.4: The raw frequencies at which outcome $a = 0$ was obtained for every pair of preparation and measurement settings. The maximum standard deviations in the data are $\sim 4 \times 10^{-3}$. Every entry in the left-most column is equal to 1—this represents the unit measurement effect which returns $a = 0$ regardless of the state of the input. The striped pattern of the data is simply an artefact of the order in which we chose to implement the preparations and measurements (described in Appendix B.2).

minimizes the weighted χ^2 statistic, defined as $\chi^2 \equiv \sum_i \sum_j \frac{\left(\tilde{D}_{ij}^{\text{realized}} - F_{ij}\right)^2}{(\Delta F_{ij})^2}$, where $(\Delta F_{ij})^2$ is the statistical variance in F_{ij}. This minimization problem is known as the weighted low-rank approximation problem, which is a non-convex optimization problem with no analytical solution [138, 139]. Nonetheless, one can use a fitting algorithm based on an alternating-least-squares method [139]. In the algorithm, it is important to constrain the entries of $\tilde{D}^{\text{realized}}$ to lie within the interval $[0, 1]$ so that they may be interpreted as probabilities. Full details are provided in Appendix B.3.

To estimate the rank of the true model underlying the data, one must compare different candidate model ranks. (For our experiment, we consider $k \in \{2, 3, \ldots, 10\}$.) For each candidate rank k, one first computes the χ^2 of the best-fit model of that rank, denoted χ^2_k,

in order to determine the extent to which each model might *underfit* the data. Second, one computes for the best-fit model of each rank the Akaike information criterion (AIC) score [140, 141] in order to determine the relative extent to which the various models *overfit* the data.

We begin by describing the method by which one finds the rank-k probability matrix $\tilde{D}^{\text{realized}}$ which minimizes χ^2 and thus best fits the data matrix F. Note that an $m \times n$ matrix with rank k is specified by a set of $r_k = k(m+n-k)$ real parameters [142], thus if the true probability matrix D^{realized} is rank k, then we expect that χ^2_k will be sampled from a χ^2 distribution with $mn - k(m+n-k) = (m-k)(n-k)$ degrees of freedom [97].

For our experiment, we calculate the variances $(\Delta F_{ij})^2$ in the expression for χ^2 by assuming that the number of detected coincident photons follows a Poissonian distribution. Fig. 5.5(a) displays the interval containing 99% of the probability density for a χ^2 distribution with $(m-k)(n-k)$ degrees of freedom, as well as χ^2_k, for each value of $k \in \{2, 3, \ldots, 10\}$. For $k < 4$, χ^2_k lies far outside the expected 99% range, and we rule out these models with high confidence.

The Kullback-Leibler (KL) divergence is a measure of the information lost when some probability distribution f is used to represent some other distribution g [143], and the AIC score of a candidate model is a measure of the Kullback-Leibler (KL) divergence between the candidate model and the true model underlying the data. Since the true model isn't known, the KL divergence can't be calculated exactly. What each candidate model's AIC score represents is its KL divergence from the true model, *relative* to all models in the candidate set. The candidate model with the lowest AIC score is closest to the true model (in the KL sense), and thus it is the most likely representation of the data among the set of candidates. These scores can be used to determine which model among a set of candidate models is the most likely to describe the data. Specifically, the likelihood can be quantified by first defining the AIC difference, Δ_k, for the rank k model as $\Delta_k = AIC_k - \min_{k'} AIC_{k'}$, then calculating the AIC weights $w_k = e^{-\frac{1}{2}\Delta_k} / \sum_{k=2}^{10} e^{-\frac{1}{2}\Delta_k}$ [143]. The AIC weight w_k represents the likelihood that the rank k model is the model that best describes the data, relative to the other models in the set of candidate models.

The AIC value for a rank-k candidate model is defined as $AIC_k = \chi^2_k + 2r_k$ [143]. The first term rewards models in proportion to how well they fit the data, and the second term penalizes models in proportion to their complexity, as measured by the number of parameters. For our experiment, the set of candidate models is the set of best-fit rank-k models for $k \in \{2, \ldots, 10\}$. We plot the AIC values for each candidate model in Fig. 5.5(b). AIC_k is minimized for $k = 4$, and we conclude that the true model underlying our dataset is

most likely rank 4. The relative likelihood of each candidate model is shown in Fig. 5.5(c). We find $w_4 = 0.9998, w_5 = 1.99 \times 10^{-4}$, and $w_k < 10^{-12}$ for other values of k.

The χ^2 goodness-of-fit test indicates that the best-fit rank-4 model fits the data well, and the AIC test indicates that the best-fit rank-4 model is the most likely of all nine candidate models to have generated the data, with relative probability 0.9998. We conclude with high confidence that the GPT that best describes our experiment has rank 4. The rank-4 matrix $\tilde{D}^{\text{realized}}$ that best fits the data provides our best estimate of the GPT state and effect vectors realized in the experiment.

5.4.5 Estimating the realized GPT state and effect spaces

The realized GPT state space, $\mathcal{S}_{\text{realized}}$ and the realized GPT state space, $\mathcal{E}_{\text{realized}}$ define the probability matrix D^{realized} from which the measurement outcomes in the experiment are sampled.

As we have described above, $\tilde{D}$realized denotes our best estimate of the true probability matrix D^{realized}. To obtain an estimate of the realized GPT state and effect spaces from $\tilde{D}^{\text{realized}}$, we must decompose it in the manner described in Sec. 3.2, that is, as $\tilde{D}^{\text{realized}} = \tilde{S}^{\text{realized}}\tilde{E}^{\text{realized}}$.

Recall that this decomposition is not unique. A convenient choice is a modified form of the singular-value decomposition, where one constrains the first column of $\tilde{S}^{\text{realized}}$ to be a column of ones, and one constrains the other three columns of $\tilde{S}^{\text{realized}}$ to be orthogonal to the first (a detailed description of this decomposition is given in Appendix B.4).

If quantum theory is the correct theory of nature, then the experimental data should be consistent with the GPT state space being the Bloch Sphere and the GPT effect space being the Bloch Diamond (depicted in Fig. 3.1(a)), up to a linear invertible transformation. Our estimate of the realized GPT state space, $\tilde{\mathcal{S}}_{\text{realized}}$, is simply the convex hull of the rows of the matrix $\tilde{S}^{\text{realized}}$. Since we have the freedom to post-process measurement outcomes, our estimate of the realized GPT effect space is slightly more complicated. There are two classes of classical post-processings that can be performed on a binary-outcome measurement. We call the first class of post-processings the *outcome-swapped class*. In this post-processing procedure, the outcome returned by a measurement device is deterministically swapped to the other outcome. The outcome-swapped outcome-0 effect for a specific measurement procedure is represented by the measurement's outcome-1 effect. We call the second class of post-processing the *trivial class*. A trivial post-processing is one in which the outcome returned by a measurement device is ignored, and deterministically replaced by an outcome $a \in \{0, 1\}$. The trivial post-processing with $a = 0$ represents the unit measurement

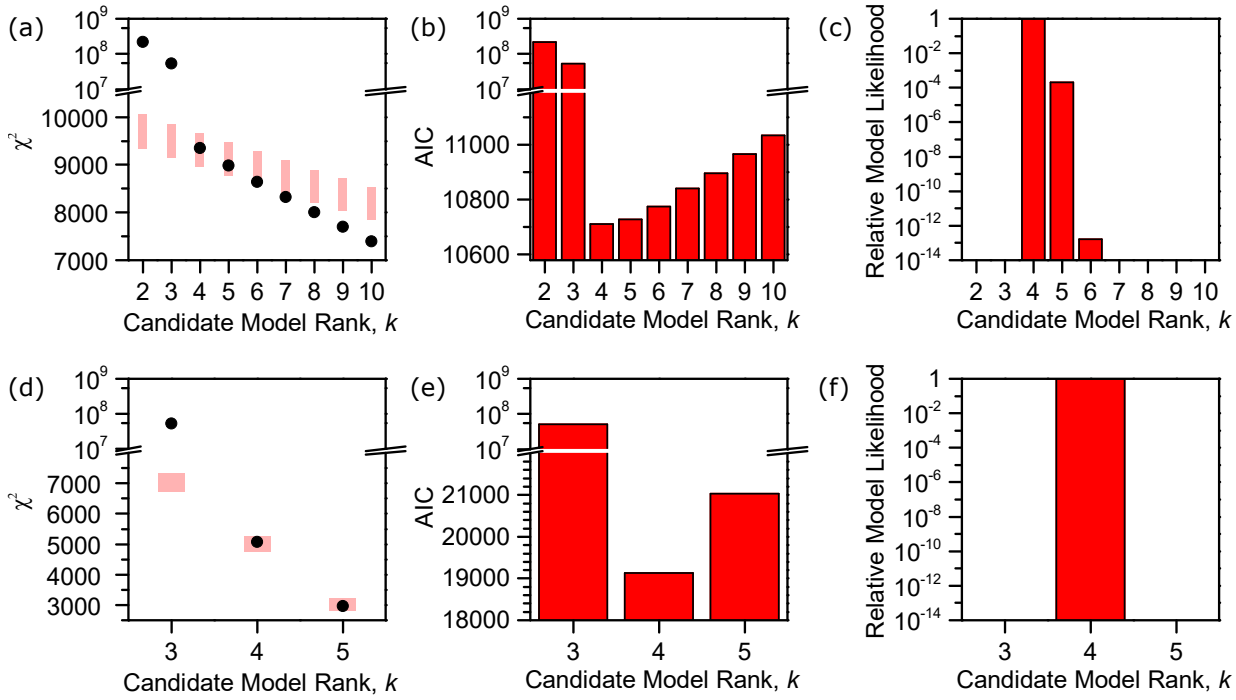

Figure 5.5: Determining the true rank of the model underlying the datasets for our two experiments. (a-c) is data for the first experiment, in which we characterized 100 preparation and measurement procedures. (d-f) is for the second experiment, in which we characterized 1006 preparation and measurement procedures. (a),(d) Comparison of the fitted χ^2 value to the expected value for a good fit, for various model ranks. Black circles are χ^2 values returned by our fitting routine. Light red bars indicate the interval in which we expect (with 99% confidence) the χ^2-value to lie, under the assumption that the true model underlying the data is rank-k. Models with $k < 4$ do not fit either dataset well. (b),(e) AIC scores for each candidate model of best fit. For both datasets the rank-4 model has the lowest AIC score, and therefore is most likely the best model among the set of candidate models. (c),(f) Relative likelihood of each model in the set of candidate models (each model without a bar has a relative likelihood less than 10^{-25}). For both datasets, the rank-4 model is most likely to describe the data.

effect, and the trivial post-processing with $a = 1$ is the outcome-swapped unit effect. Of course, one can take convex mixtures of post-processings as well, by, for example, accepting the outcome from a measurement device with probability p, and implementing a post-processing with probability $1 - p$. Hence $\tilde{\mathcal{E}}^{\text{realized}}$ is the closure under convex mixtures and classical post-processing of the vectors defined by the columns of the matrix $\tilde{E}^{\text{realized}}$.

As we have already included the unit measurement effect in $\tilde{D}^{\text{realized}}$, it is represented in $\tilde{E}^{\text{realized}}$ as well. The outcome-1 effect vector $\mathbf{e}_{[1,M]}$ for a measurement procedure M is equal to $\mathbf{u} - \mathbf{e}_{[0,M]}$, the complement of $\mathbf{e}_{[0,M]}$. Thus $\tilde{\mathcal{E}}_{\text{realized}}$ is the convex hull of the union of the set of column vectors in the matrix $\tilde{E}^{\text{realized}}$ and the set of their complements.

Our estimate of the realized GPT state space, $\tilde{\mathcal{S}}_{\text{realized}}$, and our estimate of the realized GPT effect space, $\tilde{\mathcal{E}}_{\text{realized}}$, are displayed in Fig. 5.6(a)-(c). Omitting the first column of $\tilde{S}^{\text{realized}}$ (because it contains no information), we visualize the realized GPT state space by plotting the last three entries of each row of $\tilde{S}^{\text{realized}}$ in a 3-dimensional space (the solid light blue polytope in Fig. 5.6(a)). As all four rows of $\tilde{E}^{\text{realized}}$ contain information, the realized GPT effect space is 4-dimensional, and we visualize it by plotting two 3-dimensional projections of it (the solid light green polytopes in Fig. 5.6(b),(c)). Qualitatively, $\mathcal{S}_{\text{realized}}$ is a ball-shaped polytope, and $\tilde{\mathcal{E}}_{\text{realized}}$ is a four-dimensional diamond with a ball-shaped polytope as its base. Note that they are qualitatively what one would expect if quantum theory is the correct description of nature.

Next, we compute the duals of these spaces. How this is done is described in detail in Appendix B.5. Our estimate of the set of GPT state vectors that are consistent with the realized GPT effects, $\tilde{\mathcal{S}}_{\text{consistent}} = \text{dual}(\tilde{\mathcal{E}}_{\text{realized}})$, is plotted alongside $\tilde{\mathcal{S}}_{\text{realized}}$ in Fig. 5.6(a) as a wireframe polytope. Similarly, our estimate of the set of GPT effect vectors consistent with the realized GPT states, $\tilde{\mathcal{E}}_{\text{consistent}} = \text{dual}(\tilde{\mathcal{S}}_{\text{realized}})$, is plotted as a wireframe alongside $\tilde{\mathcal{E}}_{\text{realized}}$ in Fig. 5.6(b),(c).

The smallness of the gap between $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{consistent}}$ implies that the possibilities for the true GPT are quite limited. Obviously, our results easily exclude all of the nonquantum examples of GPTs presented in Fig. 3.1.

Our results can be used to infer limits on the extent to which the true GPT might fail to satisfy the no-restriction hypothesis. One way of doing so is by bounding the volume ratio of $\mathcal{S}$ to $\mathcal{S}_{\text{logical}}$. From the discussion in Sec. 5.3, it is clear that this is upper bounded by the volume ratio of $\mathcal{S}_{\text{realized}}$ to $\mathcal{S}_{\text{consistent}}$. Given our estimates of the latter two spaces, we can compute an estimate of this ratio. We find it to be 0.9229 ± 0.0001. The error bar is the standard deviation in the volume ratio from 100 Monte Carlo simulations. We begin each simulation by simulating a set of coincidence counts. Each set of counts is found by sampling each count from a Poisson distribution with mean and variance equal

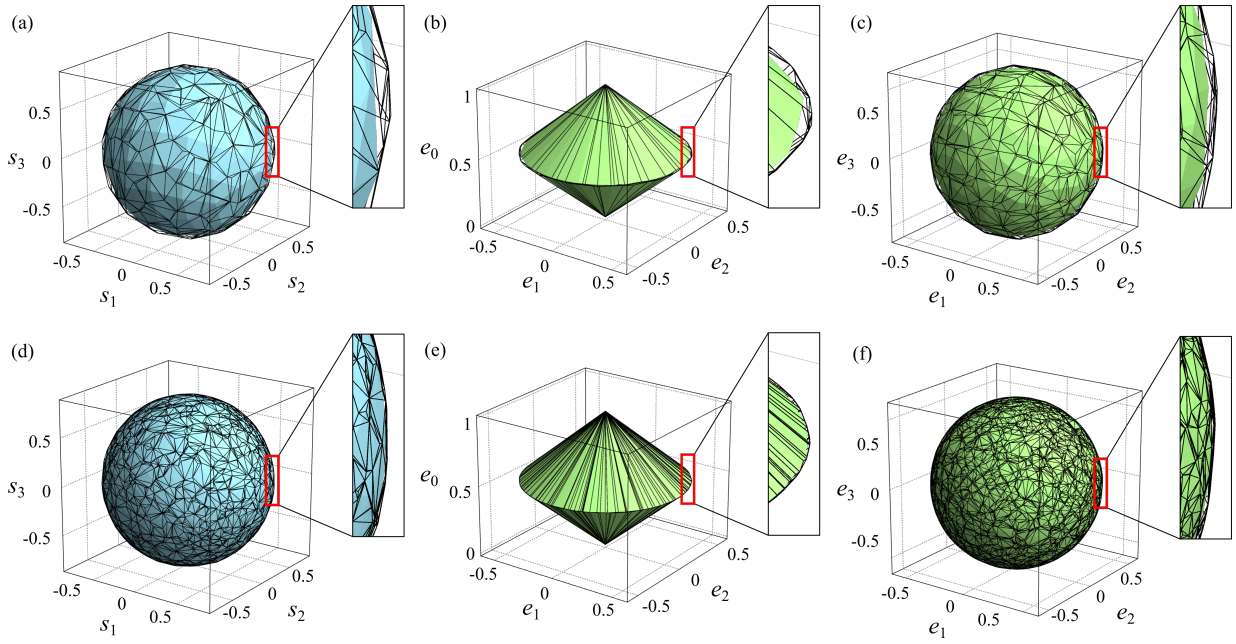

Figure 5.6: The GPT states and effects for the preparations and measurements realized in our two experiments and their duals. (a),(b),(c) First experiment, in which we characterize 100 preparation and 100 measurement procedures. (d),(e),(f) Second experiment, in which we characterize 1006 preparation and 1006 measurement procedures. (a),(d) For each experiment, the estimated space of realized GPT states, $\tilde{\mathcal{S}}_{\text{realized}}$ is the convex polytope depicted in blue, while the wireframe convex polytope which surrounds it is the estimated space of logically possible GPT states, $\tilde{\mathcal{S}}_{\text{consistent}}$, calculated from the realized GPT effects. The true state space of the GPT describing nature must lie somewhere in between $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{consistent}}$, modulo experimental uncertainty. The gap between these two spaces is smaller for the second set of data, and hence this dataset does a better job at narrowing down the possibilities for the state space. (b),(e),(c),(f) Solid green shapes are each a different 3-d projection of our estimates of the 4-d realized effect spaces, $\tilde{\mathcal{E}}_{\text{realized}}$. The wireframe convex polytopes are 3-d projections of the estimated effect space consistent with the realized preparations, $\tilde{\mathcal{E}}_{\text{consistent}}$.

to the number of photons counted in the true experiment. To our knowledge, this is the first quantitative limit on the extent to which the GPT governing nature might violate the no-restriction hypothesis.

5.4.6 Increasing the number of experimental configurations

Because the vertices of the polytopes describing $\tilde{\mathcal{S}}_{\text{realized}}$ in Figs. 5.6(a)-(c) coincide with the preparations and measurement effects that were implemented, one can safely conclude that the lack of smoothness of these polytopes is an artifact of an insufficiently dense set of experimental configurations, and not evidence for any lack of smoothness of the true GPT state and effect spaces. In order to have any chance of detecting small deviations from the smoothness predicted by quantum theory, therefore, one requires a much denser set of experimental configurations.

The lack of smoothness of the polytopes in Figs. 5.6(a)-(c) also means that the volume ratio of $\tilde{\mathcal{S}}_{\text{realized}}$ to $\tilde{\mathcal{S}}_{\text{consistent}}$ is unlikely to provide a very tight upper bound on the volume ratio of $\mathcal{S}$ to $\mathcal{S}_{\text{logical}}$. As such, having a much denser set of experimental configurations would allow one to put a stronger bound on possible deviations from the no-restriction hypothesis.

There is therefore a strong motivation to increase the number m of different preparations and the number n of different measurement effects that are probed in the experiment. It might seem at first glance that doing so is infeasible, on the grounds that it implies a significant increase in the number, mn, of preparation-measurement pairs that need to be implemented and thus an overwhelmingly long data-acquisition time.

However, this is not the case; one can probe more preparations and measurements by *not* implementing every measurement on every preparation. The key insight is that in order to characterize the GPT state vector associated to a given preparation, one needn't find its statistics on *every* measurement effect in the set being considered: it suffices to find its statistics on a subset thereof, namely, any tomographically complete subset of measurement effects. Similarly, in order to characterize the GPT effect vector associated to a given measurement effect, one need not implement it on the full set of preparations being considered, but just a tomographically complete subset thereof. The first experiment provided evidence for the conclusion that the tomographically complete sets have cardinality 4. It follows that one should be able to characterize m preparations and n measurements with just $4(m+n-4)$ experimental configurations, rather than mn.

Despite the good evidence about the cardinality from the first experiment, we deemed it worthwhile to perform the second experiment in such a manner that the analysis of the

data did not rely on any evidence drawn from the first experiment. Furthermore, we were motivated to have the second experiment provide an independent test of the hypothesis that the cardinality of the tomographically complete sets is indeed four. Given that the closest competitors to the rank-4 model on either side were those of ranks 3 and 5, we decided to restrict our set of candidate models to those having ranks in the set $k \in \{3, 4, 5\}$. In order for the experimental data to be able to reject the hypothesis of rank k as a bad fit, it is necessary that one have at least $k + 1$ measurements implemented on each preparation, and at least $k+1$ preparations on which each measurement is implemented; otherwise, one can trivially find a perfect fit. To be able to assess the quality of fit for a rank-5 model, therefore, we needed to choose at least 6 measurements that are jointly tomographically complete to implement on each of the m preparations and at least 6 preparations that are jointly tomographically complete on which each of the n measurements is implemented. We chose to use precisely 6 in each case, yielding a total of $6(m + n - 6)$ experimental configurations. Without exceeding the bound of $\sim 10^4$ experimental configurations being probed, we were able to take $m = n = 1000$ and thereby probe a factor of 10 more preparations and measurements than in the first experiment.

We refer to the set of six measurement effects (preparations) in this second experiment as the *fiducial* set. Our choice of which six waveplate settings to use in each of the fiducial sets is described in Appendix B.2. Our choice of which 1000 waveplate settings to use in order to try to densely sample the set of all preparations and measurements is also described there. (Note that although our knowledge of the quantum formalism informed both choices, our analysis of the experimental data does not presume the correctness of quantum theory.) In the end, we also implemented each of our six fiducial measurement effects on each of our six fiducial preparations, so that we had $m = n = 1006$.

We also add the unit measurement effect to our set of effects, and represent it in the first column of our data matrix. The six fiducial measurements make up the next six columns, and the six fiducial preparations make up the first six rows. We thereby arrange our data into a 1006×1007 probability matrix F, with the big difference to the first experiment being that F now has a 1000×1000 submatrix of unfilled entries.

We perform an identical analysis procedure to the one described in Sec. 5.4.4: for each k in the candidate set of ranks, we seek to find the rank-k matrix $\tilde{D}^{\text{realized}}$ of best-fit to F. For the entries in the 1000×1000 submatrix of $\tilde{D}^{\text{realized}}$ corresponding to the unfilled entries in F, the only constraint in the fit is that each entry be in the range $[0, 1]$, so that it corresponds to a probability. The results of this analysis are presented in Fig. 5.5.

The χ^2 goodness-of-fit test (Fig. 5.5(d)) rules out the rank three model, and therefore all models with rank less than 3 as well. Calculating the AIC values for the maximum-

likelihood rank 3, 4 and 5 models shows that the rank-4 model is the most likely among these to describe the data (Fig. 5.5(e),(f)). Indeed the relative probability of the rank 5 model is on the order of 10^{-414}.

The reason that the likelihood of the rank 5 model is so low is because the number of parameters required to specify a rank-k $m \times n$ matrix is $r_k = k(m + n - k)$, and since $m = n \sim 1000$, the rank 5 model requires ~ 2000 more parameters than the rank 4 model. The number of model parameters is multiplied by a factor of two in the formula for the AIC score, and the difference in χ^2_5 and χ^2_4 is only ~ 2000. This means that if the AIC score is used to calculate the likelihood of each model, the rank 5 model is $\sim e^{-2000/2} \sim 10^{-414}$ as likely as the rank 4 model.

The AIC formula we use was derived in the limit where the number of datapoints is much greater than the number of parameters in the model. In our second experiment the number of datapoints is roughly *equal* to the number of parameters in each model, and thus any conclusions which derive from use of the AIC formula must be taken with a grain of salt. We should instead use a corrected form of the AIC, called AIC_C [143]. However, the formula for AIC_C depends on the specific model being used, and to the best of our knowledge a formula has not been found for the weighted low rank approximation problem. However, every AIC_C formula that we found for different types of models *increased* the amount by which models were penalized for complexity [143]. Hence we hypothesize that the proper AIC_C formula would lead to an even smaller relative likelihood for the rank 5 model, and thus that we have strong evidence that a rank 4 model should be used to represent the second experiment. Finding the correct AIC_C formula for the weighted low rank approximation problem is an interesting problem for future consideration.

The second experiment, therefore, corroborates one of the conclusions of the first experiment, namely, that for the GPT governing single-photon polarization, the cardinality of the tomographically complete sets is four.

We decompose the rank-4 matrix of best fit and plot our estimates of the realized state space, $\tilde{\mathcal{S}}_\text{realized}$, and the realized effect space, $\tilde{\mathcal{E}}_\text{realized}$, in Fig. 5.6(d)-(f). The realized GPT state and effect spaces reconstructed from the second experiment are smoother than those from the first, and the gap between $\tilde{\mathcal{S}}_\text{realized}$ and $\tilde{\mathcal{S}}_\text{consistent}$ is smaller as well. Quantitatively, the volume ratio of $\tilde{\mathcal{S}}_\text{realized}$ to $\tilde{\mathcal{S}}_\text{consistent}$ is found to be 0.977 ± 0.001, where the error bar is calculated from 100 Monte Carlo simulations. Compared to the first experiment, this provides a tighter bound on any failure of the no-restriction hypothesis.

5.5 Bounding deviations from quantum theory in the landscape of GPTs

5.5.1 Consistency with quantum theory

We now check to see if the possibilities for the true GPT state and effect spaces implied by our analysis of the experiment include quantum theory.

As noted in Sec. 5.3, because it is in practice impossible to eliminate all noise in the experimental procedures, we expect that under the assumption that all of our realized preparations are indeed represented by quantum states, they will all be slightly impure (that is, their eigenvalues will be bounded away from 0 and 1). Their GPT state vectors should therefore be strictly in the interior of the Bloch Sphere. Similarly, we expect such noise on all of the realized measurement effects, implying that their GPT effect vectors will be strictly in the interior of the 4-dimensional Bloch Diamond. This, in turn, implies that the extremal GPT state vectors in $\mathcal{S}_{\text{consistent}}$ will be strictly in the exterior of the Bloch Sphere. The size of the gap between $\mathcal{S}_{\text{realized}}$ and $\mathcal{S}_{\text{consistent}}$, therefore, will be determined by the amount of noise in the preparations and measurements.

Naïvely, one might expect that for the quantum state and effect spaces for a qubit to be consistent with our experimental results, $\mathcal{S}_{\text{qubit}}$ must fit geometrically between our estimates of $\mathcal{S}_{\text{realized}}$ and $\mathcal{S}_{\text{consistent}}$, up to a linear transformation. That is, one might expect the condition to be that there exists a linear transformation of $\mathcal{S}_{\text{qubit}}$ that fits geometrically between $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{consistent}}$.

However, noise in the experiment also leads to statistical discrepencies between the vertices of $\tilde{\mathcal{S}}_{\text{realized}}$ and those of $\mathcal{S}_{\text{realized}}$, and between the vertices of $\tilde{\mathcal{E}}_{\text{realized}}$ and those of $\mathcal{E}_{\text{realized}}$. This noise could lead to estimates of the realized GPT state and effect vectors being longer than the actual realized GPT state and effect vectors. If the estimates of any of these lie *outside* the qubit state and effect spaces, then one could find that it is impossible to find a linear transformation of $\mathcal{S}_{\text{qubit}}$ that fits between $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{consistent}}$, even if quantum theory is correct!

We test the above intuition by simulating the first experiment under the assumption that quantum theory is the correct theory of nature. We assume that the states we actually prepare in the lab are slightly depolarized versions of the set of 100 pure quantum states that we are targeting, and that the measurements we actually perform are slightly depolarized versions of the set of 100 projective measurements we are targeting. We estimate the amount of depolarization noise from the raw data, and use the estimated amount

of noise to calculate the outcome probabilities for each depolarized measurement on each depolarized state. We arrange these probabilities into a 100×100 table and use them to simulate 1000 sets of photon counts, then analyse each of the 1000 simulated datasets with the GPT tomography procedure.

We find that, for every set of simulated data, we are unable to find a linear transformation of $\mathcal{S}_{\text{qubit}}$ that fits between the simulated $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{consistent}}$, confirming the intuition articulated above.

Nonetheless, we can quantify the closeness of the fit as follows. We find that if, for each simulation, we artificially reduce the length of the GPT vectors in the simulated $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$ by multiplying them by a factor slightly less than one, then we *can* fit a linearly transformed $\mathcal{S}_{\text{qubit}}$ between the smaller $\tilde{\mathcal{S}}_{\text{realized}}$ and larger $\tilde{\mathcal{S}}_{\text{consistent}}$. On average, we find we have to shrink the vectors making up $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$ by $0.11\% \pm 0.02\%$, where the error bar is the standard deviation over the set of simulations. To perform the above simulations we used CVX, a software package for solving convex problems [144, 145].

We quantify the real data's agreement with the simulations by performing the same calculation as on the simulated datasets. We first notice that there is no linear transformation of $\mathcal{S}_{\text{qubit}}$ that fits between $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{consistent}}$, as in the simulations. Furthermore, we find that we can achieve a fit if we shrink the vectors making up $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$ by 0.14%, which is consistent with the simulations. Thus the spaces $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$ reconstructed from the first experiment are consistent with what we expect to find given the correctness of quantum theory.

When analysing data from the second experiment it takes ~ 4 hours to run the code that solves the weighted low rank approximation problem. It is therefore impractical to perform 1000 simulations of this experiment. Instead, we extrapolate from the simulation of the first experiment.

We note two significant ways in which the second experiment differs from the first. First, we perform approximately 10 times as many preparation and measurement procedures in the second experiment than in the first, yet accumulate roughly the same amount of data. Hence, each GPT state and effect vector in the second experiment is characterized with approximately 10 times fewer detected photons than in the first experiment, and so we expect the uncertainties on the second experiment's reconstructed GPT vectors to be $\sim \sqrt{10}$ times larger than the same uncertainties in the first experiment. We expect this $\sqrt{10}$ increase in uncertainty to translate to a $\sqrt{10}$ increase in the amount we need to shrink $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$ before we can fit a linearly transformed $\mathcal{S}_{\text{qubit}}$ between $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{consistent}}$. Second, $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$ each contain 1006 GPT vectors, a factor of 10 more than in the first experiment. Since there are a greater number of GPT vectors in the

second experiment it is likely that the outliers (i.e., the cases for which our estimate differs most from the true vectors) in the second experiment will be more extreme than those in the first experiment. This should also lead to an increase in the amount we need to shrink the vectors in $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$before we can fit a linearly transformed $\mathcal{S}_{\text{qubit}}$ between $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{consistent}}$.

We find that, for the data from the second experiment, we need to shrink $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$ by 0.65%, a factor only 4 times greater than the 0.14% of the first experiment, which seems reasonable given the estimates above. We therefore conclude that the second experiment gives us no compelling reason to doubt the correctness of quantum theory.

The arguments presented above also support the notion that our experimental data is consistent with quantum theory according to the usual standards by which one judges this claim: if we had considered fitting the data with quantum states and effects rather their GPT counterparts (which one could accomplish by doing a GPT fit while *constraining* the vertices of the realized and consistent GPT state spaces to contain a sphere between them, up to linear transformations), we would have found that the quality of the fit was good.

5.5.2 Upper and lower bounds on violation of noncontextuality inequalities

One method we use to bound possible deviations from quantum theory is to consider the maximal violation of a particular noncontextuality inequality [61]. From our data we infer a range in which the maximal violation can lie, and compare this to the quantum prediction. We will briefly introduce the notion of noncontextuality, then discuss the inferences we make. The notion of noncontextuality was introduced by Kochen and Specker [40]. We here consider a generalization of the Kochen-Specker notion, termed universal noncontextuality, defined in Ref. [38].

Noncontextuality is a notion that applies to an ontological model of an operational theory. Such a model is an attempt to understand the predictions of the operational theory in terms of a system that acts as a causally mediary between the preparation device and the measurement device. It postulates a space of ontic states Λ, where the ontic state $\lambda \in \Lambda$ specifies all the physical properties of the physical system according to the model. For each preparation procedure P of a system, it is presumed that the system's ontic state λ is sampled at random from a probability distribution $p(\lambda|P)$, For each measurement M on a system, it is presumed that its outcome O is sampled at random in a manner that depends on the ontic state λ, based on the conditional probability $p(O|\lambda, M)$, It is

presumed that the empirical predictions of the operational theory are reproduced by the ontological model,

$$p(O|M,P) = \sum_{\lambda\in\Lambda} p(O|\lambda,M)p(\lambda|P). \tag{5.1}$$

We can now articulate the assumption of noncontextuality for both the preparations and the measurements.

Preparation noncontextuality. If two preparation procedures, P and P', are *operationally equivalent*, which in the GPT framework corresponds to being represented by the same GPT state vector, then they are represented by the same distribution over ontic states:

$$\mathbf{s}_P = \mathbf{s}_{P'} \implies p(\lambda|P) = p(\lambda|P'). \tag{5.2}$$

Measurement noncontextuality. If two measurement effects, $[O|M]$ and $[O'|M']$, are *operationally equivalent*, which in the GPT framework corresponds to being represented by the same GPT effect vector, then they are represented by the same distribution over ontic states:

$$\mathbf{e}_{[O|M]} = \mathbf{e}_{[O'|M']} \implies p(O|\lambda,M) = p(O'|\lambda,M'). \tag{5.3}$$

To assume *universal noncontextuality* is to assume noncontextuality for all procedures, including preparations and measurements[4].

There are now many operational inequalities for testing universal noncontextuality. Techniques for deriving such inequalities from proofs of the Kochen-Specker theorem are presented in [98, 146, 147]. In addition, there exist other proofs of the failure of universal noncontextuality that cannot be derived from the Kochen-Specker theorem. The proofs in Ref. [38] based on prepare-and-measure experiments on a single qubit are an example, and these too can be turned into inequalities testing for universal noncontextuality (as shown in Refs. [2] and [148]).

We here consider the simplest example of a noncontextuality inequality that can be violated by a qubit, namely the one associated to the task of 2-bit *parity-oblivious multiplexing* (POM), described in Ref. [61]. Bob receives as input from a referee an integer y chosen uniformly at random from $\{0,1\}$ and Alice receives a two-bit input string $(z_0, z_1) \in \{0,1\}^2$,

[4]There is also a notion of noncontextuality for transformations [38], but we will not make use of it here. In fact, the noncontextuality inequality we consider is one that only makes use of the assumption of noncontextuality for preparations.

chosen uniformly at random. Success in the task corresponds to Bob outputing the bit $b = z_y$, that is, the yth bit of Alice's input. Alice can send a system to Bob encoding information about her input, but no information about the parity of her string, $z_0 \oplus z_1$, can be transmitted to Bob. Thus, if the referee performs any measurement on the system transmitted, he should not be able to infer anything about the parity. The latter constraint is termed *parity-obliviousness*. [5]

An operational theory describes every protocol for parity-oblivious multiplexing as follows. Based on the input string $(z_0, z_1) \in \{0,1\}^2$ that she receives from the referee, Alice implements a preparation procedure $P_{z_0 z_1}$, and based on the integer $y \in \{0,1\}$ that he receives from the referee, Bob implements a binary-outcome measurement M_y, and reports the outcome b of his measurement as his output. Given that each of the 8 values of (y, z_0, z_1) are equally likely, the probability of winning, denoted $\mathcal{C}$, is

$$\mathcal{C} \equiv \frac{1}{8} \sum_{b,y,z_0,z_1} \delta_{b,z_y} p(b|P_{z_0 z_1}, M_y). \tag{5.4}$$

where δ_{b,z_y} is the Kronecker delta function. The parity obliviousness condition can be expressed as a constraint on the GPT states, as

$$\frac{1}{2}\mathbf{s}_{P_{00}} + \frac{1}{2}\mathbf{s}_{P_{11}} = \frac{1}{2}\mathbf{s}_{P_{01}} + \frac{1}{2}\mathbf{s}_{P_{10}}. \tag{5.5}$$

This asserts the operational equivalence of the parity-0 preparation (the uniform mixture of P_{00} and P_{11}) and the parity-1 preparation (the uniform mixture of P_{01} and P_{10}), and therefore it implies a nontrivial constraint on the ontological model by the assumption of preparation noncontextuality (Eq. (5.2)), namely,

$$\frac{1}{2}p(\lambda|P_{00}) + \frac{1}{2}p(\lambda|P_{11}) = \frac{1}{2}p(\lambda|P_{01}) + \frac{1}{2}p(\lambda|P_{10}). \tag{5.6}$$

It was shown in Ref. [61] that if an operational theory admits of a universally noncontextual ontological model, then the maximal value of the probability of success in parity-oblivious multiplexing is

$$\mathcal{C}_{\text{NC}} \equiv \frac{3}{4}. \tag{5.7}$$

[5] Parity-oblivious multiplexing is akin to a 2-to-1 quantum random access code. It was not introduced as a type of random access code in Ref. [61] because the latter are generally defined as having a constraint on the potential information-carrying capacity of the system transmitted, whereas in parity-oblivious multiplexing, the system can have arbitrary information-carrying capacity—the only constraint is that of parity-obliviousness.

We refer to the inequality

$$\mathcal{C} \leq \mathcal{C}_{\mathrm{NC}} \tag{5.8}$$

as the POM noncontextuality inequality. [6]

It was also shown in Ref. [61] that in operational quantum theory, the maximal value of the probability of success is

$$\mathcal{C}_{\mathrm{Q}} \equiv \frac{1}{2} + \frac{1}{2\sqrt{2}} \simeq 0.8536, \tag{5.9}$$

which violates the POM noncontextuality inequality, thereby providing a proof of the impossibility of a noncontextual model of quantum theory and demonstrating a quantum-over-noncontextual advantage for the task of parity-oblivious multiplexing. A set of four quantum states and two binary-outcome quantum measurements that satisfy the parity-obliviousness condition of Eq. (5.5) and that lead to success probability $\mathcal{C}_Q$ are illustrated in Fig. 5.8.

For a given GPT state space $\mathcal{S}$ and effect space $\mathcal{E}$, we define

$$\mathcal{C}_{(\mathcal{S},\mathcal{E})} \equiv \max_{\{\mathbf{s}_{P_{z_0 z_1}}\}\in\mathcal{S},\{\mathbf{e}_{b|M_y}\}\in\mathcal{E}} \frac{1}{8} \sum_{b,y,z_0,z_1} \delta_{b,z_y} \mathbf{s}_{P_{z_0 z_1}} \cdot \mathbf{e}_{b|M_y} \tag{5.10}$$

where the optimization must be done over choices of $\{\mathbf{s}_{P_{z_0 z_1}}\} \in \mathcal{S}$ that satisfy the parity-obliviousness constraint of Eq. (5.5). If $\mathcal{S}$ and $\mathcal{E}$ are the state and effect spaces of a GPT, then $\mathbf{s}_{P_{z_0 z_1}} \cdot \mathbf{e}_{b|M_y}$ is the probability $p(b|P_{z_0 z_1}, M_y)$ and $\mathcal{C}_{(\mathcal{S},\mathcal{E})}$ has the form of Eq. (5.4) and defines the maximum probability of success achievable in the task of parity-oblivious multiplexing for that GPT. (We will see below that it is also useful to consider $\mathcal{C}_{(\mathcal{S},\mathcal{E})}$ when the pair $\mathcal{S}$ and $\mathcal{E}$ *do not* define the state and effect spaces of a GPT.)

As discussed in Section 5.3, no experiment can specify $\mathcal{S}$ and $\mathcal{E}$ exactly. Instead, what we find is a set of possibilities for $(\mathcal{S}, \mathcal{E})$ that are consistent with the data, and thus are candidates for the true GPT state and effect spaces. We denote this set of candidates by $\mathtt{GPT}_{\mathrm{candidates}}$. To determine the range of possible values of the POM noncontextuality inequality violation in this set, we need to determine

$$\mathcal{C}_{\mathrm{min}} \equiv \min_{(\mathcal{S},\mathcal{E})\in\mathtt{GPT}_{\mathrm{candidates}}} \mathcal{C}_{(\mathcal{S},\mathcal{E})}, \tag{5.11}$$

[6] Note that an experiment test of this inequality was also reported in Ref. [61]. However, as noted in Ref. [2], the experiment of Ref. [61] did not solve the problem of inexact operational inequivalences. Although the measured deviation from exact operatonal equivalence was found to be small, there was at the time no theoretical account of how a given value of deviation should impact the degree of violation of the POM inequality. As such, it was unclear what conclusions could be drawn for the possibility of noncontextuality from the violation of the POM inequality in that experiment.

and

$$\mathcal{C}_{\max} \equiv \max_{(\mathcal{S},\mathcal{E})\in \texttt{GPT}_{\text{candidates}}} \mathcal{C}_{(\mathcal{S},\mathcal{E})}. \tag{5.12}$$

See Fig. 5.7(a) for a schematic of the relation between the various $\mathcal{C}$ quantities we consider.

$\mathcal{C}_{\min}$ and $\mathcal{C}_{\max}$ are each defined as a solution to an optimization problem. As noted in Sec. 5.3, there is a large freedom in the choice of $\mathcal{S}$ given $\mathcal{S}_{\text{realized}}$ and $\mathcal{S}_{\text{consistent}}$, and there is a large freedom in the choice of $\mathcal{E}$ for each choice of $\mathcal{S}$. Finally, for each pair $(\mathcal{S},\mathcal{E})$ in this set, one still needs to optimize over the choice of four preparations and two measurements defining the probability of success.

It turns out that the choice of $(\mathcal{S},\mathcal{E})$ that determines $\mathcal{C}_{\min}$ is easily identified. First, note that the definition in Eq. (5.10) implies the following inference

$$\mathcal{S}' \subseteq \mathcal{S},\ \mathcal{E}' \subseteq \mathcal{E} \implies \mathcal{C}_{(\mathcal{S}',\mathcal{E}')} \leq \mathcal{C}_{(\mathcal{S},\mathcal{E})}. \tag{5.13}$$

Given that $\mathcal{S}_{\text{realized}} \subseteq \mathcal{S}$ and $\mathcal{E}_{\text{realized}} \subseteq \mathcal{E}$ for all $(\mathcal{S},\mathcal{E}) \in \texttt{GPT}_{\text{candidates}}$, it follows that

$$\mathcal{C}_{(\mathcal{S}_{\text{realized}},\mathcal{E}_{\text{realized}})} \leq \mathcal{C}_{\min}. \tag{5.14}$$

And given that $(\mathcal{S}_{\text{realized}},\mathcal{E}_{\text{realized}})$ is among the GPT candidates consistent with the data, we conclude that

$$\mathcal{C}_{\min} = \mathcal{C}_{(\mathcal{S}_{\text{realized}},\mathcal{E}_{\text{realized}})}. \tag{5.15}$$

However, calculating $\mathcal{C}_{(\mathcal{S}_{\text{realized}},\mathcal{E}_{\text{realized}})}$ still requires solving the optimization problem defined in Eq. (5.10), which is computationaly difficult.

Much more tractable is the problem of determining a *lower bound* on $\mathcal{C}_{\min}$, using a simple inner approximation to $\mathcal{S}_{\text{realized}}$ and $\mathcal{E}_{\text{realized}}$. This is the approach we pursue here. We will denote this lower bound by $\texttt{LB}(\mathcal{C}_{\min})$.

Let $\mathcal{S}^w_{\text{qubit}}$ denote the image of the qubit state space $\mathcal{S}_{\text{qubit}}$ under the partially depolarizing map $\mathcal{D}_w$, defined by

$$\mathcal{D}_w(\rho) \equiv w\rho + (1-w)\frac{1}{2}\mathbb{I}, \tag{5.16}$$

with $w \in [0,1]$. Similarly, let $\mathcal{E}^{w'}_{\text{qubit}}$ denote the image of $\mathcal{E}_{\text{qubit}}$ under $\mathcal{D}_{w'}$.

Consider the 2-parameter family of GPTs defined by $\{(\mathcal{S}^w_{\text{qubit}},\mathcal{E}^{w'}_{\text{qubit}}) : w,w' \in (0,1)\}$. These correspond to quantum theory for a qubit but with noise added to the states and to the effects. Letting w_1 be the largest value of the parameter w such that $\mathcal{S}^w_{\text{qubit}} \subseteq \mathcal{S}_{\text{realized}}$

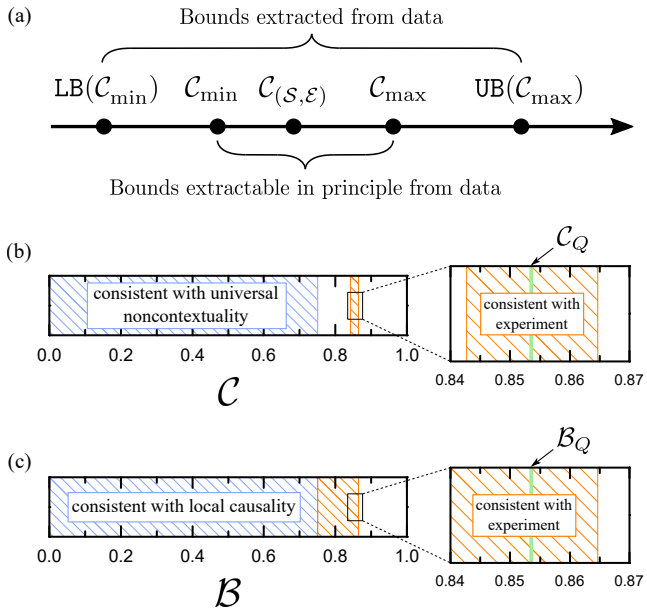

Figure 5.7: Bounding maximal inequality violations with GPT tomography. (a) Relation between the true value of the maximal violation of the POM inequality for the true GPT describing our experiment, $\mathcal{C}_{(\mathcal{S},\mathcal{E})}$, and the bounds that we place on it. The interval $[\mathcal{C}_{\min}, \mathcal{C}_{\max}]$ is the range of possible values for $\mathcal{C}_{(\mathcal{S},\mathcal{E})}$ that one can in principle infer from an experiment, and the interval $[\texttt{LB}(\mathcal{C}_{\min}), \texttt{UB}(\mathcal{C}_{\max})]$ is a conservative estimate of $[\mathcal{C}_{\min}, \mathcal{C}_{\max}]$. (b) The interval $[\texttt{LB}(\mathcal{C}_{\min}), \texttt{UB}(\mathcal{C}_{\max})]$ inferred from our data (area labelled "consistent with experiment"). The true value $\mathcal{C}_{(\mathcal{S},\mathcal{E})}$ differs from the quantum prediction, $\mathcal{C}_Q$ by at most $\pm 1.3 \pm 0.1\%$. Our data violates the POM inequality. (c) The interval $[\texttt{LB}(\mathcal{B}_{\min}), \texttt{UB}(\mathcal{B}_{\max})]$ inferred from our data (area labelled "consistent with experiment"). The true value $\mathcal{B}_{(\mathcal{S},\mathcal{E})}$ is at most $1.3 \pm 0.1\%$ greater than the maximal quantum violation, $\mathcal{C}_Q$. Error bars are too small to be visible on the plots.

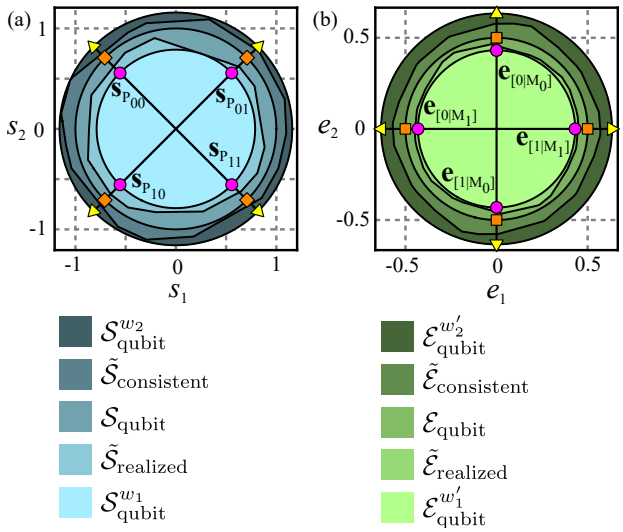

Figure 5.8: Depictions of the rescaled qubit state and effect spaces which provide our inner and outer approximations to the estimated realized GPT state and effect spaces. We also depict the states and effects that achieve the maximum probability of success in parity-oblivious multiplexing in quantum theory (orange squares), and those that achieve our lower (magenta circles) and upper (yellow triangles) bounds. The left figure depicts the GPT state vectors of the four preparations, labelled by the possible values of the pair of bits Alice must encode, and the right figure depicts the GPT effect vectors of each outcome of each of the pair of measurements.

and letting w'_1 be the largest value of the parameter w' such that $\mathcal{E}^{w}_{\text{qubit}} \subseteq \mathcal{E}_{\text{realized}}$, then $\mathcal{S}^{w_1}_{\text{qubit}}$ and $\mathcal{E}^{w'_1}_{\text{qubit}}$ provide inner approximations to $\mathcal{S}_{\text{realized}}$ and $\mathcal{E}_{\text{realized}}$ respectively, depicted in Fig. 5.8. From these, we get the lower bound

$$\texttt{LB}(\mathcal{C}_{\text{min}}) = \mathcal{C}_{(\mathcal{S}^{w_1}_{\text{qubit}}, \mathcal{E}^{w'_1}_{\text{qubit}})}. \tag{5.17}$$

A subtlety that we have avoided mentioning thus far is that the depolarized qubit state and effect spaces are only defined up to a linear transformation, so that in seeking an inner approximation, one could optimize over not only w but this linear transformation as well. To simplify the analysis, however, we took $\mathcal{S}^{w}_{\text{qubit}}$ to be a sphere of radius w and $\mathcal{E}^{w'}_{\text{qubit}}$ to be a diamond with a base that is a sphere of radius w', and we optimized over w and w'. (Optimizing over all linear transformations would simply give us a tighter lower bound.)

For the GPT $(\mathcal{S}^{w}_{\text{qubit}}, \mathcal{E}^{w'}_{\text{qubit}})$, a set of four preparations and two binary-outcome measurements that satisfy the parity-obliviousness condition of Eq. (5.5) and that yield the

maximum probability of success are the images, under the partially depolarizing maps $\mathcal{D}_w$ and $\mathcal{D}_{w'}$ respectively, of the optimal quantum choices. These images are depicted in Fig. 5.8.

One finds that with probability ww', the probability of success is the quantum value, and the rest of the time, it is just 1/2,

$$\begin{aligned}\mathcal{C}_{(\mathcal{S}^w_{\text{qubit}},\mathcal{E}^{w'}_{\text{qubit}})} &= ww'(\frac{1}{2}+\frac{1}{2\sqrt{2}})+(1-ww')\frac{1}{2},\\ &= \frac{1}{2}+ww'\frac{1}{2\sqrt{2}}.\end{aligned} \tag{5.18}$$

From our estimates of the realized GPT state and effect spaces, $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$, we obtain an estimate of w_1 by identifying the largest value of w such that $\mathcal{S}^w_{\text{qubit}} \subseteq \tilde{\mathcal{S}}_{\text{realized}}$ and we obtain an estimate of w'_1 by identifying the largest value of w' such that $\mathcal{E}^{w'}_{\text{qubit}} \subseteq \tilde{\mathcal{E}}_{\text{realized}}$.

Determining these estimates from the data of the first experiment and substituting into Eqs. (5.17) and (5.18), we infer the lower bound $\texttt{LB}(\mathcal{C}_{\text{min}}) = 0.8303 \pm 0.0002$. A similar analysis for the second experiment yields an even tighter bound,

$$\texttt{LB}(\mathcal{C}_{\text{min}}) = 0.8427 \pm 0.0005. \tag{5.19}$$

This provides a lower bound on the interval of $\mathcal{C}$ values in which the true value could be found, as depicted in Fig. 5.7(b).[7]

We now turn to $\mathcal{C}_{\text{max}}$. Given that for all $(\mathcal{S},\mathcal{E}) \in \texttt{GPT}_{\text{candidates}}$, $\mathcal{S} \subseteq \mathcal{S}_{\text{consistent}}$ and $\mathcal{E} \subseteq \mathcal{E}_{\text{consistent}}$, it follows from Eq. (5.13) that $\mathcal{C}_{\text{max}} \leq \mathcal{C}_{(\mathcal{S}_{\text{consistent}},\mathcal{E}_{\text{consistent}})}$. [8] We can therefore compute an upper bound on $\mathcal{C}_{\text{max}}$ using outer approximations to $\mathcal{S}_{\text{consistent}}$ and $\mathcal{E}_{\text{consistent}}$. We choose outer approximations consisting of rescaled qubit state and effect spaces, defined as before, but where the parameter w can now fall outside the interval $[0,1]$.

Letting w_2 be the smallest value of the parameter w such that $\mathcal{S}_{\text{consistent}} \subseteq \mathcal{S}^w_{\text{qubit}}$ and letting w'_2 be the smallest value of the parameter w' such that $\mathcal{E}_{\text{consistent}} \subseteq \mathcal{E}^{w'}_{\text{qubit}}$, then $\mathcal{S}^{w_2}_{\text{qubit}}$

[7] Note that it is likely that this lower bound could be improved if one supplemented the preparations and measurements that were implemented in the experiment with a set that were targeted towards achieving the largest value of $\mathcal{C}$ (according to quantum expectations).

[8] At this point, the analogy to the case of $\mathcal{C}_{\text{min}}$ might lead one to expect that $\mathcal{C}_{\text{max}} = \mathcal{C}_{(\mathcal{S}_{\text{consistent}},\mathcal{E}_{\text{consistent}})}$. However, this is incorrect because the pair $(\mathcal{S}_{\text{consistent}},\mathcal{E}_{\text{consistent}})$ is *not* among the GPT candidates consistent with the experimental data. In fact, it does not even correspond to a valid GPT, as one can find a GPT state vector in $\mathcal{S}_{\text{consistent}}$ and a GPT effect vector in $\mathcal{E}_{\text{consistent}}$ with inner product outside the interval $[0,1]$, hence not defining a probability. Unfortunately, if one wants to calculate $\mathcal{C}_{\text{max}}$, it seems that one must perform the difficult optimization in Eq. (5.12).

and $\mathcal{E}_{\text{qubit}}^{w_2'}$ provide outer approximations to $\mathcal{S}_{\text{consistent}}$ and $\mathcal{E}_{\text{consistent}}$ respectively, and so we get an upper bound

$$\texttt{UB}(\mathcal{C}_{\max}) = \mathcal{C}_{(\mathcal{S}_{\text{qubit}}^{w_2}, \mathcal{E}_{\text{qubit}}^{w_2'})} \tag{5.20}$$

Even though we are now allowing supernormalized state and effect vectors, via w and w' values outside of $[0, 1]$, a simple calculation shows that $\mathcal{C}_{(\mathcal{S}_{\text{qubit}}^{w}, \mathcal{E}_{\text{qubit}}^{w'})}$ is still given by Eq. (5.18).

Our estimates $\tilde{\mathcal{S}}_{\text{consistent}}$ and $\tilde{\mathcal{E}}_{\text{consistent}}$ for the state and effect spaces of the first experiment imply estimates for w_2 and w_2' [9] and substituting these into Eqs. (5.20) and (5.18), we infer $\texttt{UB}(\mathcal{C}_{\max}) = 0.8784 \pm 0.0002$. The same analysis on the second experiment yields

$$\texttt{UB}(\mathcal{C}_{\max}) = 0.8647 \pm 0.0005. \tag{5.21}$$

This provides an upper bound on the interval of $\mathcal{C}$ values in which the true value could be found, as depicted in Fig. 5.7(b).

Recalling that the quantum value is $\mathcal{C}_Q \simeq 0.8536$, it follows from Eqs. (5.19) and (5.21) that the scope for the true GPT to differ from quantum theory in the amount of contextuality it predicts (relative to the POM inequality) is quite limited: for the true GPT, the maximum violation of the POM noncontextuality inequality can be at most $1.3 \pm 0.1\%$ less than and at most $1.3 \pm 0.1\%$ more than the quantum value.

5.5.3 Upper bound on violation of Bell inequalities

Bell's theorem famously shows that a certain set of assumptions, which includes local causality, is in contradiction with the predictions of operational quantum theory [39]. It is also possible to derive inequalities from these assumptions that refer only to operational quantities and thus can be tested directly experimentally.

The Clauser, Horne, Shimony and Holt (CHSH) inequality [48] is the standard example. A pair of systems are prepared together according to a preparation procedure P^{AB}, then one is sent to Alice and the other is sent to Bob. At each wing of the experiment, the system is subjected to one of two binary-outcome measurements, M_0^A or M_1^A on Alice's side and M_0^B and M_1^B on Bob's side, with the choice of measurement being made uniformly at

[9] We note that the duality relation $\mathcal{E}_{\text{consistent}} = \text{dual}(\mathcal{S}_{\text{realized}})$ implies that $\mathcal{E}_{\text{qubit}}^{w_2'} = \text{dual}(\mathcal{S}_{\text{qubit}}^{w_1})$ and similarly, the relation $\mathcal{S}_{\text{consistent}} = \text{dual}(\mathcal{E}_{\text{realized}})$ implies $\mathcal{S}_{\text{qubit}}^{w_2} = \text{dual}(\mathcal{E}_{\text{qubit}}^{w_1'})$. This in turn implies that $w_2' = \frac{1}{w_1}$ and $w_2 = \frac{1}{w_1'}$, so that $\mathcal{C}_{\min} = \frac{1}{2} + w_1 w_1' \frac{1}{2\sqrt{2}}$ and $\mathcal{C}_{\max} = \frac{1}{2} + \frac{1}{w_1 w_1'} \frac{1}{2\sqrt{2}}$.

random, and where the choice at one wing is space-like separated from the registration of the outcome at the other wing. Denoting the binary variable determining the measurement choice at Alice's (Bob's) wing by x (y), and the outcome of Alice's (Bob's) measurement by a (b), the operational quantity of interest, the "Bell quantity" for CHSH, is defined as follows (where $a, b, x, y \in \{0, 1\}$, and $\oplus$ is addition modulo 2)

$$\mathcal{B} \equiv \frac{1}{4} \sum_{a,b,x,y} \delta_{a\oplus b,xy} p(a, b|M_x^A, M_y^B, P^{AB}). \tag{5.22}$$

The maximum value that this quantity can take in a model satisfying local causality and the other assumptions of Bell's theorem is

$$\mathcal{B}_{\text{loc}} \equiv \frac{3}{4}, \tag{5.23}$$

so that such models satisfy the CHSH inequality

$$\mathcal{B} \leq \mathcal{B}_{\text{loc}}. \tag{5.24}$$

Meanwhile, the maximum quantum value is

$$\mathcal{B}_{\text{Q}} \equiv \frac{1}{2} + \frac{1}{2\sqrt{2}} \simeq 0.8536. \tag{5.25}$$

Experimental tests have exhibited a violation of the CHSH inequality [149] and various loopholes for escaping the conclusions have been sealed experimentally [150, 151, 152, 153, 154, 155]. These experiments provide a lower bound on the value of the Bell quantity, which violates the local bound. It has not been previously clear, however, how to derive an *upper* bound on the Bell quantity. Doing so is necessary if one hopes to experimentally rule out post-quantum correlations such as the Popescu-Rohrlich box [49, 57]. We here demonstrate how to do so.

The first fact to note is that in the context of a GPT, just like in quantum theory, if Alice and Bob share a correlated bipartite state, then by implementing a measurement on her system, Alice can steer Bob's system to an ensemble of GPT states. (The bipartite state may even be entangled; in a GPT, an entangled state on a composite system is one that cannot be written as a convex mixture of states that factorize on the vector spaces of the components [122]). The state that Bob's system is steered to depends on the measurement Alice chooses to perform and the outcome she receives. If Alice performs measurement M_x^A and receives outcome a, we denote the GPT state that Bob's system is steered to by $\mathbf{s}_{a|x}^B$.

The assumption of space-like separation implies that there is no signalling between Alice and Bob, and this constrains how Bob's system can be steered. If $p(a|x) \equiv p(a|M_x^A, P^{AB})$ is the probability that Alice obtains outcome a given that she performs measurement M_x^A on the preparation P^{AB}, then the average marginal GPT state of Bob's subsystem is given by $\sum_a p(a|x)\mathbf{s}_{a|x}^B$. The no-signalling assumption forces this marginal state to be independent of Alice's measurement choice. In the CHSH scenario the no-signalling constraint is summarized with the following equation:

$$p(0|0)\mathbf{s}_{0|0}^B + p(1|0)\mathbf{s}_{1|0}^B = p(0|1)\mathbf{s}_{0|1}^B + p(1|1)\mathbf{s}_{1|1}^B. \tag{5.26}$$

Our experiment shows that it is *logically possible* for there to be bipartite states together with measurements on Alice's system which steer Bob's system to an ensemble of states *as long as* the states in this ensemble are within the GPT state space $\mathcal{S}_{\text{realized}}$. It cannot, however, attest to the *physical existence* of such bipartite states and measurements on the remote system. Because we are assuming that the true GPT includes classical probability theory as a subtheory (see Sec. 3.2), it follows that the local value is a lower limit on the range of possible values of the Bell quantity among experimentally viable candidates for the true GPT. In order to obtain a nontrivial lower limit on this range, however, one needs to perform an experiment involving two physical systems such that one can learn which GPT states for the bipartite system are physically realizable, in particular, whether there are any entangled states that are realized.

On the other hand, in spite of probing only a single system rather than a pair, our experiment *can* attest to the nonexistence of any bipartite state and measurement on Alice's system which would steer Bob's system to ensembles of states *outside* of $\mathcal{S}_{\text{consistent}}$. The reason is that if such a bipartite state and such a measurement were to exist, they could be used to prepare GPT states on Bob's system which assign values outside $[0, 1]$ (which cannot be interpreted as probabilities) to some GPT effects in $\mathcal{E}_{\text{realized}}$ on Bob's system. Therefore, we *can* use our experimental results to determine an *upper limit* on the range of values of the Bell quantity among experimentally viable candidates for the true GPT.

The maximum violation of the CHSH inequality achievable if Bob's system is described by a state space $\mathcal{S}$ and an effect space $\mathcal{E}$, is

$$\mathcal{B}_{(\mathcal{S},\mathcal{E})} \equiv \max_{\{\mathbf{s}_{P_{a|x}^B}\}\in\mathcal{S}, \{\mathbf{e}_{b|M_y^B}\}\in\mathcal{E}} \frac{1}{4} \sum_{a,b,x,y} \delta_{a\oplus b,xy} p_{a|x} \mathbf{s}_{P_{a|x}^B} \cdot \mathbf{e}_{b|M_y^B}, \tag{5.27}$$

where one varies over $\{\mathbf{s}_{P_{a|x}^B}\}$ that satisfy the no-signalling constraint, Eq. (5.26). If the

pair $\mathcal{S}$ and $\mathcal{E}$ together form a valid GPT, then $\mathbf{s}_{P^B_{a|x}} \cdot \mathbf{e}_{b|M^B_y}$ is a probability and we recover Eq. (5.22).

The upper limit on the range of possible values of the CHSH inequality violation among the theories in $\texttt{GPT}_{\text{candidates}}$, which we denote by $\mathcal{B}_{\max}$, is defined analogously to $\mathcal{C}_{\max}$ in Eq. (5.12).

Calculating $\mathcal{B}_{\max}$ is a difficult optimization problem that involves varying over every pair $(\mathcal{S}, \mathcal{E})$ consistent with the experiment, and for each pair implementing the optimization in Eq. (5.27).

Instead of performing this difficult optimization, we will derive an upper bound on $\mathcal{B}_{\max}$, denoted $\texttt{UB}(\mathcal{B}_{\max})$. This is achieved in the same manner that the upper bound on $\mathcal{C}_{\max}$ was obtained in the previous section, namely, using a qubit-like outer approximation.

For qubit-like state and effect spaces, it turns out that any upper bound on violations of the POM noncontextuality inequality implies a corresponding upper bound on violations of the CHSH inequality. This follows from the well-known fact that the optimal violation of the CHSH inequality is achieved when the marginal on Alice's outcomes is uniform. The proof is provided in Appendix B.6.

Thus, we infer from Eq. (5.21) that

$$\texttt{UB}(\mathcal{B}_{\max}) = 0.8647 \pm 0.0005. \tag{5.28}$$

This provides an upper bound on the interval of $\mathcal{B}$ values in which the true value of the maximal CHSH inequality violation lies, as depicted in Fig. 5.7(c). As noted earlier, our experiment only provides the trivial lower bound $\texttt{LB}(\mathcal{B}_{\min}) = \mathcal{B}_{\text{loc}}$. Nontrivial lower bounds have, of course, been provided in previous Bell experiments using photon polarization, such as Ref. [156].

5.6 Discussion

As we have emphasized, conclusions regarding the tomographic completeness of a given set of preparations or measurements are always tentative: they can be falsified but not confirmed. Nontheless, our first experiment provides good evidence for the conclusion that the cardinality of the tomographically complete set of preparations for photon polarization is four: of the 100 preparations we implemented, just four of these are sufficient to predict the statistics of the other 96 (for the 100 measurements considered). The same can be said

of the tomographically complete set of measurements. Our second experiment adds to this evidence.

The rank of the GPT describing our experiment can be determined with very high confidence by our method. Because the models we consider have $k(m+n-k)$ parameters, increasing the rank k of the model beyond the rank suggested by quantum theory increases the parameter count by hundreds in the first experiment and by thousands in the second. For this reason, the Akaike information criterion can deliver a decisive vote against higher-rank models on the grounds that they grossly overfit the data.

Our experimental results are consistent with the conclusion that in prepare-and-measure experiments, photon polarization acts like a 2-level quantum system. More importantly, the technique we have described provides a means of obtaining experimental bounds on possible deviations from quantum theory. We focussed in this article on two examples of such deviations, namely, the failure of the no-restriction hypothesis, and supra-quantum violations of noncontextuality and Bell inequalities.

Modifications of quantum theory that posit intrinsic decoherence imply unavoidable noise and thereby a failure of the no-restriction hypothesis. We focused on the volume ratio of $\mathcal{S}_{\text{logical}}$ to $\mathcal{S}$ as a generic measure of the failure of no restriction hypothesis, and we obtained an upper bound on that measure via the volume ratio of $\mathcal{S}_{\text{consistent}}$ to $\mathcal{S}_{\text{realized}}$. This provides an upper bound on the degree of noise in any intrinsic decoherence mechanism.

If one makes more explicit assumptions about the decoherence mechanism, one can be a bit more explicit about the bound. Suppose that the noise that arises from intrinsic decoherence in a prepare-and-measure experiment corresponds to a partially depolarizing map $\mathcal{D}_{1-\epsilon}$ (Eq. (5.16)) where ϵ is a small parameter describing the strength of the noise, then GPT tomography would find $\mathcal{S}_{\text{realized}} \subseteq \mathcal{S}^{v}_{\text{qubit}}$ and $\mathcal{E}_{\text{realized}} \subseteq \mathcal{S}^{v'}_{\text{qubit}}$ where $vv' = 1-\epsilon$. The best qubit-like inner approximations to $\mathcal{S}_{\text{realized}}$ and $\mathcal{E}_{\text{realized}}$, denoted by $\mathcal{S}^{w_1}_{\text{qubit}}$ and $\mathcal{S}^{w'_1}_{\text{qubit}}$ in our article, define a lower bound on vv', namely, $w_1 w'_1 \leq vv'$, and thereby an upper bound on ϵ, namely, $\epsilon \leq 1 - w_1 w'_1$. From our second experiment, we obtained the estimate $w_1 w'_1 = 0.969 \pm 0.001$, which implies that $\epsilon \leq 0.031 \pm 0.001$.

We have also provided experimental bounds on the amount by which the system we studied could yield Bell and noncontextuality inequality violations in excess of their maximum quantum value.

Because violation of each of the inequalities we have considered is related to an advantage for some information-processing task—specifically, parity-oblivious multiplexing and the CHSH game—it follows that our experimental upper bounds on these violations imply an upper bound on the possible advantage for these tasks. More generally, our techniques

can be used to derive limits on advantages for any task that is powered by nonlocality or contextuality.

Our results also exclude deviations from quantum theory that have some theoretical motivation. For instance, Brassard et al. [157] have shown that communication complexity becomes trivial if one has CHSH inequality violations of $\frac{1}{2} + \frac{1}{\sqrt{6}} \simeq 0.908$ or higher. If one assumes that this is the actual threshold at which communication complexitybecomes nontrivial (as opposed to being a nonstrict upper bound) and if one endorses the nontriviality of communication complexity as a principle that the true theory of the world ought to satisfy, then one has reason to speculate that the true theory of the world might achieve a CHSH inequality violation somewhere between the quantum bound of 0.8536 and 0.908. Our experimental bound, however, rules out most of this range of values.

As a test (and exclusion) of the hypothesis of universal noncontextuality, our experiment represents a significant improvement over the best previous experiment [2]. The reason is that it addresses what was identified in Ref. [2] to be the greatest weakness of that experiment, namely, the extent of the evidence for the claim that a given set of measurements or preparations should be considered tomographically complete. Recall that every assessment of operational equivalence among two preparations (measurements) rests upon the assumption that one has compared their statistics for a tomographically complete set of measurements (preparations).

The experiment reported in Ref. [2] implemented eight distinct effects and eight distinct states and proceeded to demonstrate that four effects and four states were sufficient to predict the statistics for all of the others, hence providing *some* evidence for the cardinalities of the tomographically complete sets of effects and states being four, in agreement with the quantum prediction. Such experimental evidence is, however, quite weak. As noted in Ref. [2], one can and should provide stronger evidence for any claim of tomographic completeness, because every claim of operational equivalence rests on correctly identifying the tomographically complete sets. The experiment reported herein overcomes this deficiency of the previous experiment by providing much stronger evidence in support of the claim that the set of measurements (preparations) that were implemented are in fact tomographically complete.

It is important to recall that our experiment probed only a single type of system: the polarization degree of freedom of a photon. A question that naturally arises at this point is: to what extent can our conclusions be ported to other types of systems?

Consider first the question of portability to other types of *two-level* systems (by which we mean systems which are described quantumly by a two-dimensional Hilbert space), in particular, two-level systems which we know have nontrivial interactions (for instance, by

virtue of the fact that information encoded in one can be reliably transferred to another). If it were the case that such two-level systems could be governed by different GPTs, this would immediately lead to a thorny problem of how to ensure that the different restrictions on their behaviours were respected even in the presense of interactions between them. Indeed, the principle that every n-level system has the same GPT state and effect spaces as every other has featured in many reconstructions of quantum theory within the GPT framework (see, e.g., the subspace axiom in Ref. [50], and its derivation from other axioms in Ref. [158]) and is taken to be a very natural assumption. This suggests that there are good theoretical grounds for thinking that our experimental constraints on possible deviations from quantum theory are applicable to *all* types of two-level systems.

It is less clear what conclusions one might draw for n-level systems when $n \neq 2$. For instance, although quantumly the maximum violation of a CHSH inequality is the same regardless of whether Bob's system is a qubit or a qutrit, this might not be the case for some nonquantum GPT. Therefore, although there are theoretical reasons for believing that our upper bound on the degree of CHSH inequality violation applies to all two-level systems we cannot apply those reasons to argue that violations will be bounded in this way for n-level systems. Nonetheless, if one does assume that all two-level systems are described by the same GPT, then we have constraints on the state and effect spaces of every two-level system embedded within the n-level system. This presumably restricts the possibilities for the state and effect spaces of the n-level system itself. How to infer such restrictions—for instance, how to infer an upper bound on the maximal CHSH inequality violation for a three-level system from an upper bound on the maximal CHSH inequality violation on a two-level system—is an interesting problem for future research.

There is evidently a great deal of scope for further experiments of the type described here. An obvious direction for future work is to apply our techniques to the characterization of higher dimensional systems and composites. Another interesting extension would be to generalize the technique to include GPT tomography of transformations, in addition to preparations and measurements. This is the GPT analogue of quantum process tomography, on which there has been a great deal of work due to its application in benchmarking experimental implementations of gates for quantum computation. It is likely that many ideas in this sphere can be ported to the GPT context. A particularly interesting case to consider is the scheme known as *gate set tomography* [35, 34, 36], which achieves a high-precision characterization of a set of quantum gates in a self-consistent manner.

Chapter 6

Conclusions and outlook

We presented two experimental tests of the physical theory describing nature, whatever it may be. The first experiment showed that any ontological model of nature cannot be noncontextual, and the second experiment placed bounds on the maximum amount that quantum theory can differ from the true physical theory of nature. The data from both experiments was analysed within the GPT framework, and the conclusions we made from each experiment rest on the assumption that we successfully implemented sets of preparation and measurement procedures that are tomographically complete for the GPT describing that experiment. We gathered evidence in favour of this assumption by collecting data for a large numbers of preparations and measurements and showed that a small number of measurements was sufficient to characterize each preparation, and vice versa. An interesting avenue for future work is to investigate whether or not there are other types of experiments that could provide evidence either for or against the assumption of tomographic completeness.

One difficulty that was encountered when collecting data for both experiments was that the alignment of some optical components in the setup would sometimes slowly drift over time, which introduced additional degrees of freedom to the data. As mentioned above, some obvious extensions to the experiment in Chapter 5 include applying the GPT tomography scheme to higher-dimensional systems, or to extend the scheme such that it can also characterize state transformations in the GPT framework. All of these ideas will benefit from a different experimental design with greater stability; one possibility could be to perform future experiments on a chip with integrated waveguides.

The method of GPT tomography introduced in this thesis has exciting potential for future use. When analysing data in the GPT framework, one can draw conclusions that

do not rely on any specific physical theory. Experiments of this type could then be used to gather evidence for or against competing physical theories. For example, if a specific axiom or principle of a physical theory could be defined operationally (examples might be the principles of local causality or noncontextuality, or the existance of a no-cloning theorem), then data analysed within the GPT framework could be used to confirm or rule out that principle. An interesting avenue for future research could be to identify the candidate post-quantum theories, and directly test their axioms using the GPT framework.

References

[1] 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," *arXiv:1710.05948*, 2017.

[2] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, "An experimental test of noncontextuality without unphysical idealizations," *Nature Communications*, vol. 7, p. 11780, 2016.

[3] J. M. Donohue, M. D. Mazurek, and K. J. Resch, "Theory of high-efficiency sum-frequency generation for single-photon waveform conversion," *Phys. Rev. A*, vol. 91, p. 033809, Mar 2015.

[4] X. Ma, T. Jackson, H. Zhou, J. Chen, D. Lu, M. D. Mazurek, K. A. G. Fisher, X. Peng, D. Kribs, K. J. Resch, Z. Ji, B. Zeng, and R. Laflamme, "Pure-state tomography with the expectation value of Pauli operators," *Phys. Rev. A*, vol. 93, p. 032140, Mar 2016.

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. New York, NY, USA: Cambridge University Press, 10th ed., 2011.

[6] J. J. Sakurai and J. Napolitano, *Modern quantum mechanics: second edition*. Boston: Addison-Wesley, 2011.

[7] C. C. Gerry and P. L. Knight, *Introductory quantum optics*. Cambridge, UK: Cambridge University Press, 2005.

[8] K. J. Resch, *Quantum optics: course notes*. Waterloo: University of Waterloo, 2012.

[9] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, "Dialogue concerning two views on quantum coherence: Factist and fictionist," *International Journal of Quantum Information*, vol. 04, no. 01, pp. 17–43, 2006.

[10] R. Loudon, *The quantum theory of light*. New York: Oxford University Press, 1983.

[11] P. Grangier, G. Roger, and A. Aspect, "Experimental evidence for a photon anticorrelation effect on a beam splitter: A new light on single-photon interferences," *Europhys. Lett.*, vol. 1, no. 4, p. 173, 1986.

[12] F. Bussières, J. A. Slater, N. Godbout, and W. Tittel, "Fast and simple characterization of a photon pair source," *Opt. Express*, vol. 16, pp. 17060–17069, Oct 2008.

[13] R. W. Boyd, *Nonlinear optics*. London: Academic, 3rd ed., 2008.

[14] D. Hamel, "Realization of novel entangled photon sources using periodically poled materials," Master's thesis, University of Waterloo, Waterloo, Ontario, Canada, 2010.

[15] L. Vermeyden, "Fundamental tests of quantum mechanics using two-photon entanglement," Master's thesis, University of Waterloo, Waterloo, Ontario, Canada, 2014.

[16] K. J. Resch, J. S. Lundeen, and A. M. Steinberg, "Experimental observation of nonclassical effects on single-photon detection rates," *Phys. Rev. A*, vol. 63, p. 020102, Jan 2001.

[17] J. N. Damask, *Polarization optics in telecommunications*. New York: Springer, 2005.

[18] R. Simon and N. Mukunda, "Minimal three-component SU(2) gadget for polarization optics," *Physics Letters A*, vol. 143, no. 4, pp. 165 – 169, 1990.

[19] F. D. Zela, "Two-component gadget for transforming any two nonorthogonal polarization states into one another," *Physics Letters A*, vol. 376, no. 19, pp. 1664 – 1668, 2012.

[20] Z. Hradil, "Quantum-state estimation," *Phys. Rev. A*, vol. 55, pp. R1561–R1564, Mar 1997.

[21] K. Banaszek, G. M. D'Ariano, M. G. A. Paris, and M. F. Sacchi, "Maximum-likelihood estimation of the density matrix," *Phys. Rev. A*, vol. 61, p. 010304, Dec 1999.

[22] Z. Hradil, J. Summhammer, G. Badurek, and H. Rauch, "Reconstruction of the spin state," *Phys. Rev. A*, vol. 62, p. 014101, Jun 2000.

[23] J. Řeháček, Z. Hradil, and M. Ježek, "Iterative algorithm for reconstruction of entangled states," *Phys. Rev. A*, vol. 63, p. 040303, Mar 2001.

[24] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, "Measurement of qubits," *Phys. Rev. A*, vol. 64, p. 052312, Oct 2001.

[25] I. L. Chuang and M. A. Nielsen, "Prescription for experimental determination of the dynamics of a quantum black box," *Journal of Modern Optics*, vol. 44, no. 11-12, pp. 2455–2467, 1997.

[26] J. F. Poyatos, J. I. Cirac, and P. Zoller, "Complete characterization of a quantum process: The two-bit quantum gate," *Phys. Rev. Lett.*, vol. 78, pp. 390–393, Jan 1997.

[27] J. L. O'Brien, G. J. Pryde, A. Gilchrist, D. F. V. James, N. K. Langford, T. C. Ralph, and A. G. White, "Quantum process tomography of a controlled-not gate," *Phys. Rev. Lett.*, vol. 93, p. 080502, Aug 2004.

[28] J. M. Chow, J. M. Gambetta, L. Tornberg, J. Koch, L. S. Bishop, A. A. Houck, B. R. Johnson, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Randomized benchmarking and process tomography for gate errors in a solid-state qubit," *Phys. Rev. Lett.*, vol. 102, p. 090502, Mar 2009.

[29] A. Luis and L. L. Sánchez-Soto, "Complete characterization of arbitrary quantum measurement processes," *Phys. Rev. Lett.*, vol. 83, pp. 3573–3576, Nov 1999.

[30] J. Fiurášek, "Maximum-likelihood estimation of quantum measurement," *Phys. Rev. A*, vol. 64, p. 024102, Jul 2001.

[31] G. M. D'Ariano, L. Maccone, and P. L. Presti, "Quantum calibration of measurement instrumentation," *Phys. Rev. Lett.*, vol. 93, p. 250407, Dec 2004.

[32] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, C. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio, and I. A. Walmsley, "Tomography of quantum detectors," *Nature Phys.*, vol. 5, pp. 27–30, Jan 2009.

[33] K. Fisher, *Photons & Phonons: A room-temperature diamond quantum memory.* PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 2016.

[34] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, "Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit," 2013.

[35] S. T. Merkel, J. M. Gambetta, J. A. Smolin, S. Poletto, A. D. Córcoles, B. R. Johnson, C. A. Ryan, and M. Steffen, "Self-consistent quantum process tomography," *Phys. Rev. A*, vol. 87, p. 062119, Jun 2013.

[36] D. Greenbaum, "Introduction to quantum gate set tomography," 2015.

[37] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, "Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography," *Nature Communications*, vol. 8, pp. EP –, Feb 2017. Article.

[38] R. W. Spekkens, "Contextuality for preparations, transformations, and unsharp measurements," *Phys. Rev. A*, vol. 71, p. 052108, 2005.

[39] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," *Physics*, vol. 1, no. 3, pp. 195–200, 1964.

[40] S. Kochen and E. Specker, "The problem of hidden variables in quantum mechanics," *Indiana Univ. Math. J.*, vol. 17, pp. 59–87, 1968.

[41] J. S. Bell, "On the problem of hidden variables in quantum mechanics," *Reviews of Modern Physics*, vol. 38, pp. 447–452, 1966.

[42] N. D. Mermin, "Hidden variables and the two theorems of John Bell," *Rev. Mod. Phys.*, vol. 65, pp. 803–815, Jul 1993.

[43] A. Peres, "Incompatible results of quantum measurements," *Physics Letters A*, vol. 151, no. 3, pp. 107 – 108, 1990.

[44] N. D. Mermin, "Simple unified form for the major no-hidden-variables theorems," *Phys. Rev. Lett.*, vol. 65, pp. 3373–3376, Dec 1990.

[45] A. Peres, "Two simple proofs of the Kochen-Specker theorem," *Journal of Physics A: Mathematical and General*, vol. 24, no. 4, p. L175, 1991.

[46] A. Cabello, J. Estebaranz, and G. García-Alcaine, "Bell-Kochen-Specker theorem: A proof with 18 vectors," *Physics Letters A*, vol. 212, pp. 183–187, Mar. 1996.

[47] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, "Simple test for hidden variables in spin-1 systems," *Phys. Rev. Lett.*, vol. 101, p. 020403, Jul 2008.

[48] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," *Phys. Rev. Lett.*, vol. 23, pp. 880–884, Oct 1969.

[49] B. S. Tsirelson, "Quantum generalizations of Bell's inequality," *Letters in Mathematical Physics*, vol. 4, no. 2, pp. 93–100, 1980.

[50] L. Hardy, "Quantum theory from five reasonable axioms. Preprint at http://arxiv.org/abs/quant-ph/0101012 (2001).." 2001.

[51] J. Barrett, "Information processing in generalized probabilistic theories," *Phys. Rev. A*, vol. 75, p. 032304, Mar 2007.

[52] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Probabilistic theories with purification," *Phys. Rev. A*, vol. 81, p. 062348, Jun 2010.

[53] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Informational derivation of quantum theory," *Phys. Rev. A*, vol. 84, p. 012311, Jul 2011.

[54] L. Hardy, "Foliable operational structures for general probabilistic theories," in *Deep Beauty: Understanding the quantum world through mathematical innovation* (H. Halvorson, ed.), ch. 11, pp. 409–442, Cambridge: Cambridge University Press, 2009.

[55] B. Schumacher and M. D. Westmoreland, *Almost Quantum Theory*, pp. 45–81. Dordrecht: Springer Netherlands, 2016.

[56] A. J. Short and J. Barrett, "Strong nonlocality: a trade-off between states and measurements," *New Journal of Physics*, vol. 12, no. 3, p. 033034, 2010.

[57] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," *Foundations of Physics*, vol. 24, no. 3, pp. 379–385, 1994.

[58] R. W. Spekkens, "Evidence for the epistemic view of quantum states: A toy theory," *Physical Review A*, vol. 75, no. 3, p. 032110, 2007.

[59] R. W. Spekkens, "Negativity and contextuality are equivalent notions of nonclassicality," *Phys. Rev. Lett.*, vol. 101, no. 2, p. 20401, 2008.

[60] C. Ferrie and J. Emerson, "Frame representations of quantum mechanics and the necessity of negativity in quasi-probability representations," *J. Phys. A: Math. Theor.*, vol. 41, no. 35, p. 352001, 2008.

[61] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, "Preparation contextuality powers parity-oblivious multiplexing," *Phys. Rev. Lett.*, vol. 102, p. 010401, 2009.

[62] M. Howard, J. Wallman, V. Veitch, and J. Emerson, "Contextuality supplies the 'magic' for quantum computation.," *Nature*, vol. 510, pp. 351–5, June 2014.

[63] R. Raussendorf, "Contextuality in measurement-based quantum computation," *Phys. Rev. A*, vol. 88, p. 022322, Aug 2013.

[64] M. J. Hoban, E. T. Campbell, K. Loukopoulos, and D. E. Browne, "Non-adaptive measurement-based quantum computation and multi-party Bell inequalities," *New J. Phys.*, vol. 13, no. 2, p. 023014, 2011.

[65] M. F. Pusey, "Anomalous weak values are proofs of contextuality," *Phys. Rev. Lett.*, vol. 113, p. 200401, Nov. 2014.

[66] D. A. Meyer, "Finite precision measurement nullifies the Kochen-Specker theorem," *Phys. Rev. Lett.*, vol. 83, pp. 3751–3754, Nov 1999.

[67] A. Kent, "Noncontextual hidden variables and physical measurements," *Phys. Rev. Lett.*, vol. 83, pp. 3755–3757, Nov 1999.

[68] R. Clifton and A. Kent, "Simulating quantum mechanics by non-contextual hidden variables," *Proc. R. Soc. Lond. A*, vol. 456, no. 2001, pp. 2101–2114, 2000.

[69] N. D. Mermin, "A Kochen-Specker theorem for imprecisely specified measurement." 1999.

[70] C. Simon, Č. Brukner, and A. Zeilinger, "Hidden-variable theorems for real experiments," *Phys. Rev. Lett.*, vol. 86, pp. 4427–4430, May 2001.

[71] J.-Å. Larsson, "A Kochen-Specker inequality," *Europhys. Lett.*, vol. 58, pp. 799–805, 2002.

[72] J. Barrett and A. Kent, "Non-contextuality, finite precision measurement and the Kochen-Specker theorem," *Stud. Hist. Phil. Mod. Phys.*, vol. 35, no. 2, pp. 151 – 176, 2004.

[73] A. Cabello and G. Garcia-Alcaine, "Proposed experimental tests of the Bell-Kochen-Specker theorem," *Phys. Rev. Lett.*, vol. 80, p. 1797, Mar. 1998.

[74] A. Cabello, "Kochen-Specker theorem and experimental test on hidden variables," *Int. J. Mod. Phys. A*, vol. 15, no. 18, pp. 2813–2820, 2000.

[75] C. Simon, M. Żukowski, H. Weinfurter, and A. Zeilinger, "Feasible "Kochen-Specker" experiment with single particles," *Phys. Rev. Lett.*, vol. 85, pp. 1783–1786, Aug 2000.

[76] A. Cabello, S. Filipp, H. Rauch, and Y. Hasegawa, "Proposed experiment for testing quantum contextuality with neutrons," *Phys. Rev. Lett.*, vol. 100, p. 130404, Apr 2008.

[77] A. Cabello, "Experimentally testable state-independent quantum contextuality," *Phys. Rev. Lett.*, vol. 101, p. 210401, Nov 2008.

[78] P. Badziag, I. Bengtsson, A. Cabello, and I. Pitowsky, "Universality of state-independent violation of correlation inequalities for noncontextual theories," *Phys. Rev. Lett.*, vol. 103, p. 050401, Jul 2009.

[79] O. Gühne, M. Kleinmann, A. Cabello, J.-A. Larsson, G. Kirchmair, F. Zähringer, R. Gerritsma, and C. F. Roos, "Compatibility and noncontextuality for sequential measurements," *Phys. Rev. A*, vol. 81, p. 022121, Feb 2010.

[80] M. Michler, H. Weinfurter, and M. Żukowski, "Experiments towards falsification of noncontextual hidden variable theories," *Phys. Rev. Lett.*, vol. 84, pp. 5457–5461, Jun 2000.

[81] Y.-F. Huang, C.-F. Li, Y.-S. Zhang, J.-W. Pan, and G.-C. Guo, "Experimental test of the Kochen-Specker theorem with single photons," *Phys. Rev. Lett.*, vol. 90, p. 250401, Jun 2003.

[82] Y. Hasegawa, R. Loidl, G. Badurek, M. Baron, and H. Rauch, "Quantum contextuality in a single-neutron optical experiment," *Phys. Rev. Lett.*, vol. 97, p. 230401, Dec 2006.

[83] G. Kirchmair, F. Zahringer, R. Gerritsma, M. Kleinmann, O. Guhne, A. Cabello, R. Blatt, and C. F. Roos, "State-independent experimental test of quantum contextuality," *Nature*, vol. 460, pp. 494–497, Jul 2009.

[84] H. Bartosik, J. Klepp, C. Schmitzer, S. Sponar, A. Cabello, H. Rauch, and Y. Hasegawa, "Experimental test of quantum contextuality in neutron interferometry," *Phys. Rev. Lett.*, vol. 103, p. 040403, Jul 2009.

[85] E. Amselem, M. Rådmark, M. Bourennane, and A. Cabello, "State-independent quantum contextuality with single photons," *Phys. Rev. Lett.*, vol. 103, p. 160405, Oct 2009.

[86] O. Moussa, C. A. Ryan, D. G. Cory, and R. Laflamme, "Testing contextuality on quantum ensembles with one clean qubit," *Phys. Rev. Lett.*, vol. 104, p. 160501, Apr 2010.

[87] R. Lapkiewicz, P. Li, C. Schaeff, N. K. Langford, S. Ramelow, M. Wiesniak, and A. Zeilinger, "Experimental non-classicality of an indivisible quantum system," *Nature*, vol. 474, pp. 490–493, Jun 2011.

[88] R. W. Spekkens, "The status of determinism in proofs of the impossibility of a noncontextual model of quantum theory," *Found. Phys.*, vol. 44, no. 11, pp. 1125–1155, 2014.

[89] A. Winter, "What does an experimental test of quantum contextuality prove or disprove?," *J. Phys. A: Math. Theor.*, vol. 47, no. 42, p. 424031, 2014.

[90] J. V. Kujala, E. N. Dzhafarov, and J.-Å. Larsson, "Necessary and sufficient conditions for contextuality in a broad class of quantum mechanical systems." 2014.

[91] C. Stark, "Self-consistent tomography of the state-measurement gram matrix," *Phys. Rev. A*, vol. 89, p. 052109, May 2014.

[92] C. J. Stark and A. W. Harrow, "Compressibility of positive semidefinite factorizations and quantum models." 2014.

[93] T. Kim, M. Fiorentino, and F. N. C. Wong, "Phase-stable source of polarization-entangled photons using a polarization Sagnac interferometer," *Phys. Rev. A*, vol. 73, p. 012316, Jan 2006.

[94] A. Fedrizzi, T. Herbst, A. Poppe, T. Jennewein, and A. Zeilinger, "A wavelength-tunable fiber-coupled source of narrowband entangled photons," *Opt. Express*, vol. 15, pp. 15377–15386, Nov 2007.

[95] D. N. Biggerstaff, R. Kaltenbaek, D. R. Hamel, G. Weihs, T. Rudolph, and K. J. Resch, "Cluster-state quantum computing enhanced by high-fidelity generalized measurements," *Phys. Rev. Lett.*, vol. 103, p. 240504, Dec 2009.

[96] M. Krystek and M. Anton, "A weighted total least-squares algorithm for fitting a straight line," *Meas. Sci. Technol.*, vol. 18, no. 11, p. 3438, 2007.

[97] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press, 3 ed., 2007.

[98] R. Kunjwal and R. W. Spekkens, "From the Kochen-Specker theorem to noncontextuality inequalities without assuming determinism," *Phys. Rev. Lett.*, vol. 115, p. 110403, Sep 2015.

[99] M. F. Pusey, "The robust noncontextuality inequalities in the simplest scenario. Preprint at arxiv:1506.04178 (2015)..".

[100] G. Chiribella and R. W. Spekkens, *Introduction*, pp. 1–18. Dordrecht: Springer Netherlands, 2016.

[101] G. C. Ghirardi, A. Rimini, and T. Weber, "Unified dynamics for microscopic and macroscopic systems," *Phys. Rev. D*, vol. 34, pp. 470–491, Jul 1986.

[102] I. C. Percival, "Primary state diffusion," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 447, no. 1929, pp. 189–209, 1994.

[103] G. J. Milburn, "Intrinsic decoherence in quantum mechanics," *Phys. Rev. A*, vol. 44, p. 5401, 1991.

[104] S. L. Adler, "Remarks on a proposed super-kamiokande test for quantum gravity induced decoherence effects," *Phys. Rev. D*, vol. 62, p. 117901, Oct 2000.

[105] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, "Almost quantum correlations," *Nature Communications*, vol. 6, p. 6288, 2015.

[106] A. B. Sainz, Y. Guryanova, A. Acín, and M. Navascués, "Almost quantum correlations violate the no-restriction hypothesis." 2017.

[107] R. D. Sorkin, "Quantum mechanics as quantum measure theory," *Modern Physics Letters A*, vol. 09, no. 33, pp. 3119–3127, 1994.

[108] U. Sinha, C. Couteau, T. Jennewein, R. Laflamme, and G. Weihs, "Ruling out multi-order interference in quantum mechanics," *Science*, vol. 329, no. 5990, pp. 418–421, 2010.

[109] J. M. Hickmann, E. J. S. Fonseca, and A. J. Jesus-Silva, "Born's rule and the interference of photons with orbital angular momentum by a triangular slit," *EPL (Europhysics Letters)*, vol. 96, no. 6, p. 64006, 2011.

[110] I. Söllner, B. Gschösser, P. Mai, B. Pressl, Z. Vörös, and G. Weihs, "Testing Born's rule in quantum mechanics for three mutually exclusive events," *Foundations of Physics*, vol. 42, pp. 742–751, Jun 2012.

[111] D. K. Park, O. Moussa, and R. Laflamme, "Three path interference using nuclear magnetic resonance: a test of the consistency of Born's rule," *New Journal of Physics*, vol. 14, no. 11, p. 113025, 2012.

[112] T. Kauten, R. Keil1, T. Kaufmann, B. Press, Časlav Brukner, and G. Weihs, "Obtaining tight bounds on higher-order interferences with a 5-path interferometer," *New Journal of Physics*, vol. 19, p. 033017, 2017.

[113] A. Peres, "Proposed test for complex versus quaternion quantum theory," *Phys. Rev. Lett.*, vol. 42, pp. 683–686, Mar 1979.

[114] S. L. Adler, "Generalized quantum dynamics," *Nuclear Physics B*, vol. 415, no. 1, pp. 195 – 242, 1994.

[115] S. L. Adler, *Quaternionic quantum mechanics and quantum fields*. New York: Oxford University Press, 1995.

[116] H. Barnum, M. Graydon, and A. Wilce, "Composites and categories of euclidean jordan algebras," *arXiv preprint arXiv:1606.09331*, 2016.

[117] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge: Cambridge University Press, 2006.

[118] B. Dakić and v. Brukner, "Quantum theory and beyond: Is entanglement special?," in *Deep Beauty: Understanding the quantum world through mathematical innovation* (H. Halvorson, ed.), pp. 365–392, Cambridge: Cambridge University Press, 2009.

[119] G. M. D'Ariano, "Probabilistic theories: What is special about quantum mechanics?," in *Philosophy of Quantum Information and Entanglement* (A. Bokulich and G. Jaeger, eds.), ch. 5, pp. 85–126, Cambridge: Cambridge University Press, 2010.

[120] L. Masanes and M. P. Mller, "A derivation of quantum theory from physical requirements," *New Journal of Physics*, vol. 13, no. 6, p. 063001, 2011.

[121] P. Janotta and H. Hinrichsen, "Generalized probability theories: what determines the structure of quantum theory?," *Journal of Physics A: Mathematical and Theoretical*, vol. 47, no. 32, p. 323001, 2014.

[122] H. Barnum and A. Wilce, *Post-Classical Probability Theory*, pp. 367–420. Dordrecht: Springer Netherlands, 2016.

[123] G. W. Mackey, *The Mathematical foundations of quantum mechanics: a lecture-note volume*. New York, NY, USA: W. A. Benjamin, 1963.

[124] G. Ludwig, *Die Grundlagen der Quantenmechanik*. Berlin, Heidelberg: Springer-Verlag, 1954.

[125] G. Ludwig, "The problem: An axiomatic basis for quantum mechanics," in *Foundations of Quantum Mechanics I*, pp. 1–11, Springer, 1983.

[126] K. Kraus, *States, effects and operations: fundamental notions of quantum theory*. Springer, 1983.

[127] H. Barnum, M. P. Müller, and C. Ududec, "Higher-order interference and single-system postulates characterizing quantum theory," *New Journal of Physics*, vol. 16, no. 12, p. 123029, 2014.

[128] B. Dakić, T. Paterek, and Č Brukner, "Density cubes and higher-order interference theories," *New Journal of Physics*, vol. 16, no. 2, p. 023028, 2014.

[129] C. M. Lee and J. H. Selby, "Higher-order interference in extensions of quantum theory," *Foundations of Physics*, vol. 47, pp. 89–112, Jan 2017.

[130] K. Vogel and H. Risken, "Determination of quasiprobability distributions in terms of probability distributions for the rotated quadrature phase," *Phys. Rev. A*, vol. 40, pp. 2847–2849, Sep 1989.

[131] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, "Measurement of the wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum," *Phys. Rev. Lett.*, vol. 70, pp. 1244–1247, Mar 1993.

[132] H. Haffner, W. Hansel, C. F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Korber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Guhn e, W. Dur, and R. Blatt, "Scalable multiparticle entanglement of trapped ions," *Nature*, vol. 438, pp. 643–646, 2005.

[133] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer, R. Ozeri, R. Reichle, and D. J. Wineland, "Creation of a six-atom /'schrodinger cat/' state," *Nature*, vol. 438, pp. 639–642, 2005.

[134] T. J. Dunn, I. A. Walmsley, and S. Mukamel, "Experimental determination of the quantum-mechanical state of a molecular vibrational mode using fluorescence tomography," *Phys. Rev. Lett.*, vol. 74, pp. 884–887, Feb 1995.

[135] A. I. Lvovsky and M. G. Raymer, "Continuous-variable optical quantum-state tomography," *Rev. Mod. Phys.*, vol. 81, pp. 299–332, Mar 2009.

[136] K. R. Popper, *The logic of scientific discovery*. New York: Basic Books, 1961.

[137] X. Feng and Z. Zhang, "The rank of a random matrix," *Appl. Math. Comput.*, vol. 185, no. 1, pp. 689 – 694, 2007.

[138] N. Gillis and F. Glineur, "Low-rank matrix approximation with weights or missing data is NP-hard," *SIAM Journal on Matrix Analysis and Applications*, vol. 32, no. 4, pp. 1149–1165, 2011.

[139] I. Markovsky, *Low Rank Approximation: Algorithms, Implementation, Applications*. Springer-Verlag London, 2012.

[140] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proceedings of the Second International Symposium on Information Theory* (B. Petrov and F. Caski, eds.), pp. 267–281, ACM Press, 1972.

[141] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, pp. 716–723, Dec 1974.

[142] E. J. Candes and T. Tao, "The power of convex relaxation: Near-optimal matrix completion," *IEEE Transactions on Information Theory*, vol. 56, pp. 2053–2080, May 2010.

[143] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag New York, 2002.

[144] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1." http://cvxr.com/cvx, Mar. 2014.

[145] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control* (V. Blondel, S. Boyd, and H. Kimura, eds.), Lecture Notes in Control and Information Sciences, pp. 95–110, Springer-Verlag Limited, 2008.

[146] A. Krishna, R. W. Spekkens, and E. Wolfe, "Deriving robust noncontextuality inequalities from algebraic proofs of the Kochen-Specker theorem: the Peres-Mermin square," *arXiv preprint arXiv:1704.01153*, 2017.

[147] R. Kunjwal, "From statistical proofs of the Kochen-Specker theorem to noise-robust noncontextuality inequalities." 2017.

[148] D. Schmid and R. W. Spekkens, "Contextual advantage for state discrimination." 2017.

[149] A. Aspect, J. Dalibard, and G. Roger, "Experimental test of Bell's inequalities using time-varying analyzers," *Phys. Rev. Lett.*, vol. 49, pp. 1804–1807, Dec 1982.

[150] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, "Violation of Bell's inequality under strict Einstein locality conditions," *Phys. Rev. Lett.*, vol. 81, pp. 5039–5043, Dec 1998.

[151] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano1, C. Monroe, and D. J. Wineland, "Experimental violation of a Bell's inequality with efficient detection," *Nature*, vol. 409, pp. 791–794, 2001.

[152] C. Erven, E. Meyer-Scott, K. Fisher, J. Lavoie, B. L. Higgins, Z. Yan, C. J. Pugh, J.-P. Bourgoin, R. Prevede, L. K. Shalm, L. Richards, N. Gigov, R. Laflamm, G. Weihs, T. Jennewein, and K. J. Resch, "Experimental three-photon quantum nonlocality under strict locality conditions," *Nature Photonics*, vol. 8, pp. 292–296, 2014.

[153] B. Hensen, H. Bernien, A. E. Dreau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellan, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres," *Nature*, vol. 526, pp. 682–686, 10 2015.

[154] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, "Significant-loophole-free test of Bell's theorem with entangled photons," *Phys. Rev. Lett.*, vol. 115, p. 250401, Dec 2015.

[155] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, "Strong loophole-free test of local realism," *Phys. Rev. Lett.*, vol. 115, p. 250402, Dec 2015.

[156] B. G. Christensen, Y.-C. Liang, N. Brunner, N. Gisin, and P. G. Kwiat, "Exploring the limits of quantum nonlocality with entangled photons," *Phys. Rev. X*, vol. 5, p. 041052, Dec 2015.

[157] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, "Limit on nonlocality in any world in which communication complexity is not trivial," *Phys. Rev. Lett.*, vol. 96, p. 250401, Jun 2006.

[158] L. Hardy, *Reconstructing Quantum Theory*, pp. 223–248. Dordrecht: Springer Netherlands, 2016.

[159] R. D. Gill, "Accardi contra bell (cum mundi): The impossible coupling," *Lecture Notes-Monograph Series*, vol. 42, pp. 133–154, 2003.

[160] Y. Zhang, S. Glancy, and E. Knill, "Asymptotically optimal data analysis for rejecting local realism," *Phys. Rev. A*, vol. 84, p. 062118, Dec 2011.

[161] Y. Zhang, S. Glancy, and E. Knill, "Efficient quantification of experimental evidence against local realism," *Phys. Rev. A*, vol. 88, p. 052119, Nov 2013.

[162] P. Bierhorst, "A rigorous analysis of the Clauser-Horne-Shimony-Holt inequality experiment when trials need not be independent," *Foundations of Physics*, vol. 44, no. 7, pp. 736–761, 2014.

[163] P. Bierhorst, "A robust mathematical model for a loophole-free Clauser-Horne experiment," *Journal of Physics A: Mathematical and Theoretical*, vol. 48, no. 19, p. 195302, 2015.

[164] J. Kofler, M. Giustina, J.-Å. Larsson, and M. W. Mitchell, "Requirements for a loophole-free photonic Bell test using imperfect setting generators. Preprint at http://arxiv.org/abs/1411.4787 (2015).," 2015.

[165] E. A. Rakhmanov, E. B. Saff, and Y. M. Zhou, "Minimal discrete energy on the sphere," *Mathematical Research Letters*, vol. 1, pp. 647–662, 1994.

[166] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[167] D. C. Lay, *Linear Algebra and Its Applications (3rd Edition)*. Boston, MA, USA: Addison Wesley, 2002.

[168] D. Avis and K. Fukuda, "A pivoting algorithm for convex hulls and vertex enumeration of arrangements and polyhedra," *Discrete & Computational Geometry*, vol. 8, pp. 295–313, Sep 1992.

[169] H. Audren, "pyparma." http://pypi.python.org/pypi/pyparma, 2016.

[170] K. Fukuda, "cddlib." http://www.inf.ethz.ch/personal/fukudak/cdd_home/, 2005.

[171] S. Popescu and D. Rohrlich, "Which states violate Bell's inequality maximally?," *Physics Letters A*, vol. 169, no. 6, pp. 411 – 414, 1992.

APPENDICES

Appendix A

Appendices for Chapter 4

Notes and acknowledgements

This chapter contains the supplementary information of work that has been published as [2]:

M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch and R. W. Spekkens, "An experimental test of noncontextuality without unphysical idealizations" *Nature Communications*, **7**:11780 (2016).

Author contributions

All authors contributed to writing the supplementary information of this work.

A.1 Derivation and tightness of the bound in our noncontextuality inequality

A.1.1 Derivation of bound

In Chapter 4, we only provided an argument for why our two applications of the assumption of noncontextuality, Eqs. (4.3) and (4.5), implied that the quantity A must be bounded away from 1. Here we show that the explicit value of this bound is $\frac{5}{6}$.

By definition,

$$A \equiv \frac{1}{6} \sum_{t \in \{1,2,3\}} \sum_{b \in \{0,1\}} p(X = b | M_t, P_{t,b}). \tag{A.1}$$

Substituting for $p(X{=}b|M_t, P_{t,b})$ the expression in terms of the distribution $\mu(\lambda|P_{t,b})$ and the response function $\xi(X = b|M_t, \lambda)$ given in Eq. (4.1), we have

$$A = \frac{1}{6} \sum_{t \in \{1,2,3\}} \sum_{b \in \{0,1\}} \sum_{\lambda \in \Lambda} \xi(X = b|M_t, \lambda)\mu(\lambda|P_{t,b}). \tag{A.2}$$

We now simply note that there is an upper bound on each response function that is independent of the value of b, namely,

$$\xi(X = b|M_t, \lambda) \leq \eta(M_t, \lambda), \tag{A.3}$$

where

$$\eta(M_t, \lambda) \equiv \max_{b' \in \{0,1\}} \xi(X = b'|M_t, \lambda). \tag{A.4}$$

We therefore have

$$A \leq \frac{1}{3} \sum_{t \in \{1,2,3\}} \sum_{\lambda \in \Lambda} \eta(M_t, \lambda) \left(\frac{1}{2} \sum_{b \in \{0,1\}} \mu(\lambda|P_{t,b}) \right), \tag{A.5}$$

Recalling that P_t is an equal mixture of $P_{t,0}$ and $P_{t,1}$, so that

$$\mu(\lambda|P_t) = \frac{1}{2}\mu(\lambda|P_{t,0}) + \frac{1}{2}\mu(\lambda|P_{t,1}), \tag{A.6}$$

we can rewrite the bound as simply

$$A \leq \frac{1}{3} \sum_{t \in \{1,2,3\}} \sum_{\lambda \in \Lambda} \eta(M_t, \lambda)\mu(\lambda|P_t). \tag{A.7}$$

But recalling Eq. (4.5),

$$\forall \lambda \in \Lambda : \mu(\lambda|P_1) = \mu(\lambda|P_2) = \mu(\lambda|P_3), \tag{A.8}$$

we see that the distribution $\mu(\lambda|P_t)$ is independent of t, so we denote it by $\nu(\lambda)$ and rewrite the bound as

$$A \leq \sum_{\lambda\in\Lambda} \left(\frac{1}{3} \sum_{t\in\{1,2,3\}} \eta(M_t, \lambda) \right) \nu(\lambda). \tag{A.9}$$

This last step is the first use of noncontextuality in the proof because Eq. (A.8) is derived from preparation noncontextuality and the operational equivalence of Eq. (4.4). It then follows that

$$A \leq \max_{\lambda\in\Lambda} \left(\frac{1}{3} \sum_{t\in\{1,2,3\}} \eta(M_t, \lambda) \right). \tag{A.10}$$

Therefore, if we can provide a nontrivial upper bound on $\frac{1}{3}\sum_t \eta(M_t, \lambda)$ for an arbitrary ontic state λ, we obtain a nontrivial upper bound on A. We infer constraints on the possibilities for the triple $(\eta(M_1, \lambda), \eta(M_2, \lambda), \eta(M_3, \lambda))$ from constraints on the possibilities for the triple $(\xi(X{=}0|M_1, \lambda), \xi(X{=}0|M_2, \lambda), \xi(X{=}0|M_3, \lambda))$.

The latter triple is constrained by Eq. (4.7), which in the case of $X = 0$ reads

$$\frac{1}{3} \sum_{t\in\{1,2,3\}} \xi(X{=}0|M_t, \lambda) = \frac{1}{2}. \tag{A.11}$$

This is the second use of noncontextuality in our proof, because Eq. (A.11) is derived from the operational equivalence of Eq. (4.7) and the assumption of measurement noncontextuality.

The fact that the range of each response function is $[0, 1]$ implies that the vector $(\xi(X{=}0|M_1, \lambda), \xi(X{=}0|M_2, \lambda), \xi(X{=}0|M_3, \lambda))$ is constrained to the unit cube. The linear constraint of Eq. (A.11) implies that these vectors are confined to a two-dimensional plane. The intersection of the plane and the cube defines the polygon depicted in Fig. A.1. The six vertices of this polygon have coordinates that are a permutation of $(1, \frac{1}{2}, 0)$. For every λ, the vector $(\xi(X{=}0|M_1, \lambda), \xi(X{=}0|M_2, \lambda), \xi(X{=}0|M_3, \lambda))$ corresponds to a point in the convex hull of these extreme points and given that $\frac{1}{3}\sum_t \eta(M_t, \lambda)$ is a convex function of this vector, it suffices to find a bound on the value of this function at the extreme points. If λ is the extreme point $(1, \frac{1}{2}, 0)$, then we have $(\eta(M_1, \lambda), \eta(M_2, \lambda), \eta(M_3, \lambda)) = (1, \frac{1}{2}, 1)$, and the other extreme points are simply permutations thereof. It follows that

$$\frac{1}{3} \sum_t \eta(M_t, \lambda) \leq \frac{5}{6}. \tag{A.12}$$

Substituting this bound into Eq. (A.10), we have our result.

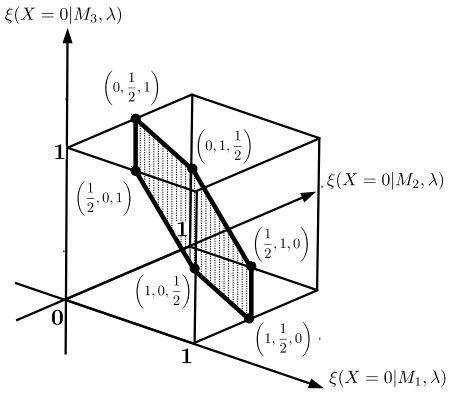

Figure A.1: Possible values of noncontextual measurement response functions. Possible values of the vector of response functions $(\xi(X{=}0|M_1,\lambda), \xi(X{=}0|M_2,\lambda), \xi(X{=}0|M_3,\lambda))$. The range of each response function is $[0,1]$, constraining the vectors to lie inside the unit cube. This fact, along with the linear constraint in Eq. (A.11), constrains the vectors to the shaded polygon.

A.1.2 Tightness of bound: an ontological model

In this section, we provide an explicit example of a noncontextual ontological model that saturates our noncontextuality inequality, thus proving that the noncontextuality inequality is tight, i.e., the upper bound of the inequality cannot be reduced any further for a noncontextual model.

For the ontological models we present, we begin by specifying the ontic state space Λ, as depicted in Fig. A.2 as a pie chart with each slice corresponding to a different element of Λ. We specify the six preparations $P_{t,b}$ by the distributions over Λ that they correspond to, denoted $\mu(\lambda|P_{t,b})$ (middle left of Fig. A.2). We specify the three measurements M_t by the response functions for the $X = 0$ outcome, denoted $\xi(0|M_t,\lambda)$ (top right of Fig. A.2). Finally, we compute the operational probabilities for the various preparation-measurement pairs, using Eq. (4.1), and display the results in the 6×4 upper-left-hand corner of Table A.1.

In the remainder of the table, we display the operational probabilities for the effective preparations, P_t, which are computed from the operational probabilities for the $P_{t,b}$ and the fact that P_t is the uniform mixture of $P_{t,0}$ and $P_{t,1}$. We also display the operational probabilities for the effective measurement M_*, which is computed from the operational

	$[0\|M_1]$	$[0\|M_2]$	$[0\|M_3]$	$[0\|M_*]$
$P_{1,0}$	5/6	1/3	1/3	1/2
$P_{1,1}$	1/6	2/3	2/3	1/2
$P_{2,0}$	1/3	5/6	1/3	1/2
$P_{2,1}$	2/3	1/6	2/3	1/2
$P_{3,0}$	1/3	1/3	5/6	1/2
$P_{3,1}$	2/3	2/3	1/6	1/2
P_1	1/2	1/2	1/2	1/2
P_2	1/2	1/2	1/2	1/2
P_3	1/2	1/2	1/2	1/2

Table A.1: Operational statistics of a noncontextual ontological model that saturates our inequality. Operational statistics from the noncontextual ontological model of Fig. A.2, achieving $A = 5/6$. The shaded cells correspond to the ones relevant for calculating A.

probabilities for the M_t and the fact that M_* is a uniform mixture of M_1, M_2 and M_3.

From Table A.1, we can verify that our ontological model implies the operational equivalences that we use in the derivation of our noncontextuality inequality. Specifically, the three preparations P_1, P_2 and P_3 yield exactly the same statistics for all of the measurements, and the measurement M_* is indistinguishable from a fair coin flip for all the preparations.

Figure A.2 also depicts $\mu(\lambda|P_t)$ for $t \in \{1, 2, 3\}$ (bottom left). These are determined from the $\mu(\lambda|P_{t,b})$ via Eq. (A.6). Similarly, the response function $\xi(0|M_*, \lambda)$, which is determined from $\xi(X = b|M_*, \lambda) = \frac{1}{3}\sum_{t\in\{1,2,3\}} \xi(X = b|M_t, \lambda)$, is displayed in the bottom right.

Given the operational equivalence of P_1, P_2 and P_3, an ontological model is preparation noncontextual if and only if $\mu(\lambda|P_1) = \mu(\lambda|P_2) = \mu(\lambda|P_3)$ for all $\lambda \in \Lambda$. We see, therefore, that the model is preparation noncontextual.

Similarly given the operational equivalence of M_* and a fair coin flip, an ontological model is measurement noncontextual if and only if $\xi(0|M_*, \lambda) = \frac{1}{2}$ for all $\lambda \in \Lambda$. We see, therefore, that the model is measurement noncontextual.

Finally, using the operational probabilities in the Table A.1, one can compute the value of A for the model. It is determined entirely by the operational probabilities in the shaded cells. One thereby confirms that $A = \frac{5}{6}$ in the noncontextual ontological model.

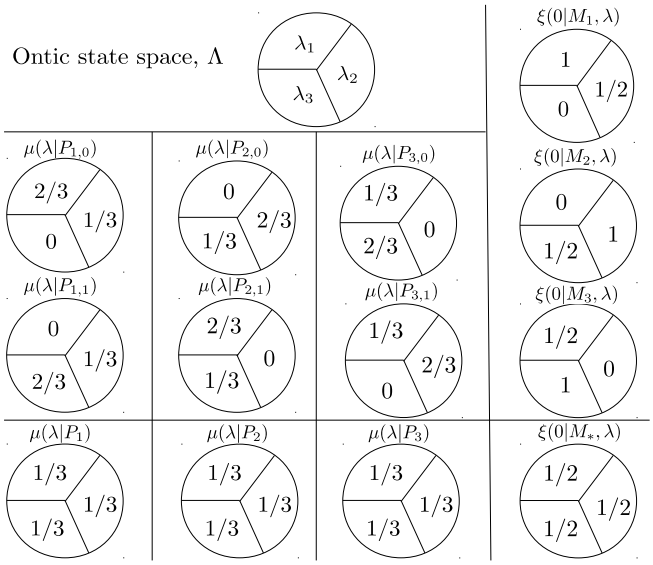

Figure A.2: Demonstration that the bound of our noncontextuality inequality is tight. The figure displays a noncontextual ontological model that saturates the noncontextual bound of our inequality.

A.2 Constructing the secondary procedures from the primary ones

A.2.1 Secondary preparations in quantum theory

As noted in Section 4.3.3, it is easiest to describe the details of our procedure for defining secondary preparations if we make the assumption that quantum theory correctly describes the experiment. Further on, we will describe the procedure for a generalised probabilistic theory (GPT).

Figure 4.1 described how to define the secondary preparations if the primary preparations deviate from the ideal only *within* the $\boldsymbol{x} - \boldsymbol{z}$ plane of the Bloch sphere. Here, we consider the case where the six primary preparations deviate from the ideals within the bulk of the Bloch sphere. The fact that our proof only requires that the secondary preparations satisfy Eq. (4.10) means that the different pairs, $P^{\mathrm{s}}_{t,0}$ and $P^{\mathrm{s}}_{t,1}$ for $t \in \{1, 2, 3\}$, need not all mix to the center of the Bloch sphere, but only to the *same* state. It follows that the three pairs need not be coplanar in the Bloch sphere. Note, however, for any *two* values, t and t', the four preparations $P^{\mathrm{s}}_{t,0}, P^{\mathrm{s}}_{t,1}, P^{\mathrm{s}}_{t',0}, P^{\mathrm{s}}_{t',1}$ do need to be coplanar.

Any mixing procedure defines a map from each of the primary preparations $P^{\mathrm{p}}_{t,b}$ to the corresponding secondary preparation $P^{\mathrm{s}}_{t,b}$, which can be visualized as a motion of the corresponding point within the Bloch sphere. To ensure that the six secondary preparations approximate well the ideal preparations while also defining mixed preparations P^{s}_1, P^{s}_2 and P^{s}_3 that satisfy the appropriate operational equivalences, the mixing procedure must allow for motion in the $\pm\boldsymbol{y}$ direction. Consider what happens if one tries to achieve such motion *without* supplementing the primary set with the eigenstates of $\boldsymbol{\sigma}\cdot\boldsymbol{y}$. A given point that is biased towards $-\boldsymbol{y}$ can be moved in the $+\boldsymbol{y}$ direction by mixing it with another point that has less bias in the $-\boldsymbol{y}$ direction. However, because the primary preparations are widely separated within the $\boldsymbol{x}-\boldsymbol{z}$ plane, achieving a small motion in $+\boldsymbol{y}$ direction in this fashion comes at the price of a large motion within the $\boldsymbol{x}-\boldsymbol{z}$ plane, implying a significant motion away from the ideal. This problem is particularly pronounced if the primary points are very close to coplanar.

The best way to move a given point in the $\pm\boldsymbol{y}$ direction is to mix it with a point that is at roughly the same location within the $\boldsymbol{x}-\boldsymbol{z}$ plane, but displaced in the $\pm\boldsymbol{y}$ direction. This scheme, however, would require supplementing the primary set with one or two additional preparations for every one of its elements. Supplementing the original set with just the two eigenstates of $\boldsymbol{\sigma}\cdot\boldsymbol{y}$ constitutes a good compromise between keeping the number of preparations low and ensuring that the secondary preparations are close to the ideal. Because the $\boldsymbol{\sigma}\cdot\boldsymbol{y}$ eigenstates have the greatest possible distance from the $\boldsymbol{x}-\boldsymbol{z}$ plane, they can be used to move any point close to that plane in the $\pm\boldsymbol{y}$ direction while generating only a modest motion within the $\boldsymbol{x}-\boldsymbol{z}$ plane.

A.2.2 Secondary measurements in quantum theory

Just as with the case of preparations, we solve the problem of no strict statistical equivalences for measurements by noting that from the primary set of measurements, M^{p}_1, M^{p}_2 and M^{p}_3, one can infer the statistics of a large family of measurements, and one can find three measurements within this family, called the secondary measurements and denoted M^{s}_1, M^{s}_2 and M^{s}_3, such that their mixture, M^{s}_*, satisfies the operational equivalence of Eq. (4.2) *exactly*. To give the details of our approach, it is again useful to begin with the quantum description.

A geometric visualization of the construction is also possible in this case. Just as a density operator can be written $\rho=\frac{1}{2}(\mathbb{I}+\boldsymbol{r}\cdot\boldsymbol{\sigma})$ to define a three-dimensional Bloch vector $\boldsymbol{r}$, an effect can be written $E=\frac{1}{2}(e_0\mathbb{I}+\boldsymbol{e}\cdot\boldsymbol{\sigma})$ to define a four-dimensional Bloch-like vector $(e_0,\boldsymbol{e})$, whose four components we will call the $\mathbb{I}$, $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ components. Note

that $e_0 = \mathrm{tr}(E)$, while $e_x = \mathrm{tr}(\boldsymbol{\sigma} \cdot \boldsymbol{x} E)$ and so forth. The eigenvalues of E are expressed in terms of these components as $\frac{1}{2}(e_o \pm |\boldsymbol{e}|)$. Consequently, the constraint that $0 \leq E \leq \mathbb{I}$ takes the form of three inequalities $0 \leq e_o \leq 2$, $|\boldsymbol{e}| \leq e_0$ and $|\boldsymbol{e}| \leq 2 - e_0$. This corresponds to the intersection of two cones. For the case $e_y = 0$, the Bloch representation of the effect space is three-dimensional and is displayed in Fig. A.3. When portraying binary-outcome measurements associated to a POVM $\{E, \mathbb{I} - E\}$ in this representation, it is sufficient to portray the Bloch-like vector $(e_0, \boldsymbol{e})$ for outcome E alone, given that the vector for $\mathbb{I} - E$ is simply $(2 - e_0, -\boldsymbol{e})$. Similarly, to describe any mixture of two such POVMs, it is sufficient to describe the mixture of the effects corresponding to the first outcome.

The family of measurements that is defined in terms of the primary set is slightly different than what we had for preparations. The reason is that each primary measurement on its own generates a family of measurements by probabilistic post-processing of its outcome. If we denote the outcome of the original measurement by X and that of the processed measurement by X', then the probabilistic processing is a conditional probability $p(X'|X)$. It is sufficient to determine the convexly-extremal post-processings, since all others can be obtained from these by mixing. For the case of binary outcome measurements considered here, there are just four extremal post-processings: the identity process, $p(X'|X) = \delta_{X',X}$; the process that flips the outcome, $p(X'|X) = \delta_{X',X\oplus 1}$; the process that always generates the outcome $X' = 0$, $p(X'|X) = \delta_{X',0}$; and the process that always generates the outcome $X' = 1$, $p(X'|X) = \delta_{X',1}$. Applying these to our three primary measurements, we obtain eight measurements in all: the two that generate a fixed outcome, the three originals, and the three originals with the outcome flipped. If the set of primary measurements corresponded to the ideal set, then the eight extremal post-processings would correspond to the observables $0, \mathbb{I}, \boldsymbol{\sigma} \cdot \boldsymbol{n_1}, -\boldsymbol{\sigma} \cdot \boldsymbol{n_1}, \boldsymbol{\sigma} \cdot \boldsymbol{n_2}, -\boldsymbol{\sigma} \cdot \boldsymbol{n_2}, \boldsymbol{\sigma} \cdot \boldsymbol{n_3}, -\boldsymbol{\sigma} \cdot \boldsymbol{n_3}$. In practice, the last six measurements will be unsharp. These eight measurements can then be mixed probabilistically to define the family of measurements from which the secondary measurements must be chosen. We refer to this family as the *convex hull of the post-processings* of the primary set.

We will again start with a simplified example, wherein the primary measurements have Bloch-like vectors with vanishing component along $\boldsymbol{y}$, $e_y = 0$, and unit component along $\mathbb{I}$, $e_0 = 1$, so that $E = \frac{1}{2}(\mathbb{I} + e_x \boldsymbol{\sigma} \cdot \boldsymbol{x} + e_z \boldsymbol{\sigma} \cdot \boldsymbol{z})$. In this case, the constraint $0 \leq E \leq \mathbb{I}$ reduces to $|\boldsymbol{e}| \leq 1$, which is the same constraint that applies to density operators confined to the $\boldsymbol{x} - \boldsymbol{z}$ plane of the Bloch sphere. Here, the only deviation from the ideal is within this plane, and the construction is precisely analogous to what is depicted in Fig. 4.1.

Unlike the case of preparations, however, the primary measurements can deviate from the ideal in the $\mathbb{I}$ direction, that is, E may have a component along $\mathbb{I}$ that deviates from 1, which corresponds to introducing a state-independent bias on the outcome of the mea-

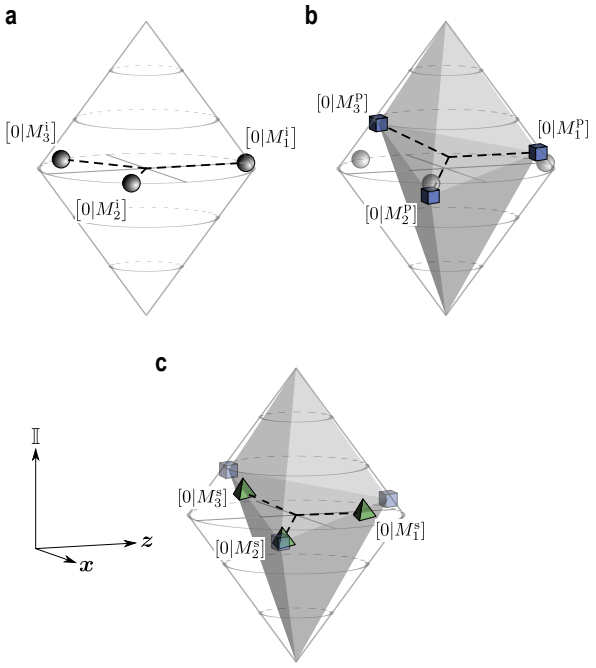

Figure A.3: Enforcing operational equivalence for measurements. A depiction of the construction of secondary measurements from primary ones in the simplified case where the component along $\boldsymbol{y}$ is zero. For each measurement, we specify the point corresponding to the Bloch representation of its first outcome. These are labelled $[0|M_1]$, $[0|M_2]$ and $[0|M_3]$. The equal mixture of these three, labelled $[0|M_*]$, is the centroid of these three points, i.e. the point equidistant from all three. **a**, The ideal measurements $[0|M_t^{\mathrm{i}}]$ with centroid at $\mathbb{I}/2$, illustrating that the operational equivalence of Eq. (4.2) is satisfied exactly. **b**, Errors in the experiment (exaggerated) will imply that the realized measurements $[0|M_t^{\mathrm{P}}]$ (termed primary) will deviate from the ideal, and their centroid deviates from $\mathbb{I}/2$. The family of points corresponding to probabilistic mixtures of the $[0|M_t^{\mathrm{P}}]$ and the observables 0 and $\mathbb{I}$ are depicted by the grey region. (For clarity, we have not depicted the outcome-flipped versions of the three primary measurements, and have not included them in the probabilistic mixtures. As we note in the text, such a restriction still allows for a good construction.) **c**, The secondary measurements M_t^{s} that have been chosen from this grey region. They are chosen such that their centroid is at $\mathbb{I}/2$, restoring the operational equivalence of Eq. (4.2).

surement. This is where the extremal post-processings yielding the constant-outcome measurements corresponding to the observables 0 and $\mathbb{I}$ come in. They allow one to move in the $\pm\mathbb{I}$ direction.

Figure A.3 presents an example wherein the primary measurements have Bloch-like vectors that deviate from the ideal not only within the $\boldsymbol{x}-\boldsymbol{z}$ plane, but in the $\mathbb{I}$ direction as well (it is still presumed, however, that all components in the $\boldsymbol{y}$ direction are vanishing).

In practice, of course, the $\boldsymbol{y}$ component of our measurements never vanishes precisely either. We therefore apply the same trick as we did for the preparations. We supplement the set of primary measurements with an additional measurement, denoted M_4^{p}, that ideally corresponds to the observable $\boldsymbol{\sigma}\cdot\boldsymbol{y}$. The post-processing which flips the outcome then corresponds to the observable $-\boldsymbol{\sigma}\cdot\boldsymbol{y}$. Mixing the primary measurements with M_4^{p} and its outcome-flipped counterpart allows motion in the $\pm\boldsymbol{y}$ direction within the Bloch cone.

Note that the capacity to move in both the $+\boldsymbol{y}$ and the $-\boldsymbol{y}$ direction is critical for achieving the operational equivalence of Eq. (4.2), because if the secondary measurements had a common bias in the $\boldsymbol{y}$ direction, they could not mix to the POVM $\{\mathbb{I}/2,\mathbb{I}/2\}$ as Eq. (4.9) requires. For the preparations, by contrast, supplementing the primary set by just *one* of the eigenstates of $\boldsymbol{\sigma}\cdot\boldsymbol{y}$ would still work, given that the mixed preparations P_t^{s} do not need to coincide with the completely mixed state $\mathbb{I}/2$.

The secondary measurements M_1^{s}, M_2^{s} and M_3^{s} are then chosen from the convex hull of the post-processings of the $M_1^{\mathrm{p}}, M_2^{\mathrm{p}}, M_3^{\mathrm{p}}, M_4^{\mathrm{p}}$. Without this supplementation, it may be impossible to find secondary measurements that define an M_*^{s} that satisfies the operational equivalences while providing a good approximation to the ideal measurements.

In all, under the extremal post-processings of the supplemented set of primary measurements, we obtain ten points which ideally correspond to the observables $0, \mathbb{I}, \boldsymbol{\sigma}\cdot\boldsymbol{n_1}, -\boldsymbol{\sigma}\cdot\boldsymbol{n_1}, \boldsymbol{\sigma}\cdot\boldsymbol{n_2}, -\boldsymbol{\sigma}\cdot\boldsymbol{n_2}, \boldsymbol{\sigma}\cdot\boldsymbol{n_3}, -\boldsymbol{\sigma}\cdot\boldsymbol{n_3}, \boldsymbol{\sigma}\cdot\boldsymbol{y}$, and $-\boldsymbol{\sigma}\cdot\boldsymbol{y}$.

Note that the outcome-flipped versions of the three primary measurements are not critical for defining a good set of secondary measurements, and indeed we find that we can dispense with them and still obtain good results. This is illustrated in the example of Fig. A.3.

A.2.3 Secondary preparations and measurements in generalised probabilistic theories

We do not want to presuppose that our experiment is well fit by a quantum description. Therefore instead of working with density operators and POVMs, we work with GPT states

and effects, which are inferred from the matrix D^{p}

$$D^{\mathrm{p}} = \begin{pmatrix} p^1_{1,0} & p^1_{1,1} & \cdots & p^1_{4,0} & p^1_{4,1} \\ p^2_{1,0} & p^2_{1,1} & \cdots & p^2_{4,0} & p^2_{4,1} \\ p^3_{1,0} & p^3_{1,1} & \cdots & p^3_{4,0} & p^3_{4,1} \\ p^4_{1,0} & p^4_{1,1} & \cdots & p^4_{4,0} & p^4_{4,1} \end{pmatrix}. \tag{A.13}$$

where

$$p^{t'}_{t,b} \equiv p(0|M^{\mathrm{p}}_{t'}, P^{\mathrm{p}}_{t,b}) \tag{A.14}$$

is the probability of obtaining outcome 0 in the t'th measurement that was actually realized in the experiment (recall that we term this measurement primary and denote it by $M^{\mathrm{p}}_{t'}$), when it follows the (t,b)th preparation that was actually realized in the experiment (recall that we term this preparation primary and denote it by $P^{\mathrm{p}}_{t,b}$). These probabilities are estimated by fitting the raw experimental data (which are merely finite samples of the true probabilities) to a GPT; we postpone the description of this procedure to Section A.3.

The rows of the D^{p} matrix define the GPT effects. We denote the vector defined by the tth row, which is associated to the measurement event $[0|M^{\mathrm{p}}_t]$ (obtaining the 0 outcome in the primary measurement M^{p}_t), by $\mathbf{M}^{\mathrm{p}}_t$. Similarly, the columns of this matrix define the GPT states. We denote the vector associated to the (t,b)th column, which is associated to the primary preparation $P^{\mathrm{p}}_{t,b}$, by $\mathbf{P}^{\mathrm{p}}_{t,b}$.

As described in Section 4.3.3, we define the *secondary* preparation $P^{\mathrm{s}}_{t,b}$ by a probabilistic mixture of the primary preparations. Thus, the GPT state of the secondary preparation is a vector $\mathbf{P}^{\mathrm{s}}_{t,b}$ that is a probabilistic mixture of the $\mathbf{P}^{\mathrm{p}}_{t,b}$,

$$\mathbf{P}^{\mathrm{s}}_{t,b} = \sum_{t'=1}^{4} \sum_{b'=0}^{1} u^{t,b}_{t',b'} \mathbf{P}^{\mathrm{p}}_{t',b'}, \tag{A.15}$$

where the $u^{t,b}_{t',b'}$ are the weights in the mixture.

A secondary measurement $M^{\mathrm{s}}_{t'}$ is obtained from the primary measurements in a similar fashion, but in addition to probabilistic mixtures, one must allow certain post-processings of the measurements, in analogy to the quantum case described above.

The set of all post-processings of the primary outcome-0 measurement events has extremal elements consisting of the outcome-0 measurement events themselves together with: the measurement event that *always* occurs, which is represented by the vector of probabilities where every entry is 1, denoted $\mathbf{1}$; the measurement event that *never* occurs (so

that outcome 1 is certain instead), which is represented by the vector of probabilities where every entry is 0, denoted $\mathbf{0}$; and the outcome-flipped measurement events, which are represented by the vector $\mathbf{1} - \mathbf{M}^{\mathrm{p}}_t$.

We can therefore define our three secondary outcome-0 measurement events as probabilistic mixtures of the four primary ones as well as the extremal post-processings mentioned above, that is

$$\mathbf{M}^{\mathrm{s}}_t = \sum_{t'=1}^{4} v^t_{t'} \mathbf{M}^{\mathrm{p}}_{t'} + v^t_{\mathbf{0}} \mathbf{0} + v^t_{\mathbf{1}} \mathbf{1} + \sum_{t''=1}^{4} v^t_{\neg t''} (\mathbf{1} - \mathbf{M}^{\mathrm{p}}_{t''}), \tag{A.16}$$

where for each t, the vector of weights in the mixture is $(v^t_1, v^t_2, v^t_3, v^t_4, v^t_{\mathbf{0}}, v^t_{\mathbf{1}}, v^t_{\neg 1}, v^t_{\neg 2}, v^t_{\neg 3}, v^t_{\neg 4})$. We see that this is a particular type of linear transformation on the rows.

Again, as mentioned in the discussion of the quantum case, we can in fact limit the post-processing to exclude the outcome-flipped measurement events for M_1, M_2 and M_3, keeping only the outcome-flipped event for M_4, and still obtain good results. Thus we found it sufficient to search for secondary outcome-0 measurement events among those of the form

$$\mathbf{M}^{\mathrm{s}}_t = \sum_{t'=1}^{4} v^t_{t'} \mathbf{M}^{\mathrm{p}}_{t'} + v^t_{\mathbf{0}} \mathbf{0} + v^t_{\mathbf{1}} \mathbf{1} + v^t_{\neg 4} (\mathbf{1} - \mathbf{M}^{\mathrm{p}}_4), \tag{A.17}$$

where for each t, the vector of weights in the mixture is $(v^t_1, v^t_2, v^t_3, v^t_4, v^t_{\mathbf{0}}, v^t_{\mathbf{1}}, v^t_{\neg 4})$.

Returning to the preparations, we choose the weights $u^{t,b}_{t',b'}$ to maximize the function

$$C_{\mathrm{P}} \equiv \frac{1}{6} \sum_{t=1}^{3} \sum_{b=0}^{1} u^{t,b}_{t,b} \tag{A.18}$$

subject to the linear constraint

$$\frac{1}{2} \sum_b \mathbf{P}^{\mathrm{s}}_{1,b} = \frac{1}{2} \sum_b \mathbf{P}^{\mathrm{s}}_{2,b} = \frac{1}{2} \sum_b \mathbf{P}^{\mathrm{s}}_{3,b}, \tag{A.19}$$

as noted in Section 4.3.4. This optimization ensures that the secondary preparations are as close as possible to the primary ones while ensuring that they satisfy the relevant operational equivalence *exactly*. Table A.2 reports the weights $u^{t,b}_{t',b'}$ that were obtained from this optimization procedure, averaged over the 100 runs of the experiment. These weights yield $C_{\mathrm{P}} = 0.9969 \pm 0.0001$, indicating that the secondary preparations are indeed very close to the primary ones.

	$P^{\mathrm{p}}_{1,0}$	$P^{\mathrm{p}}_{1,1}$	$P^{\mathrm{p}}_{2,0}$	$P^{\mathrm{p}}_{2,1}$	$P^{\mathrm{p}}_{3,0}$	$P^{\mathrm{p}}_{3,1}$	$P^{\mathrm{p}}_{4,0}$	$P^{\mathrm{p}}_{4,1}$
$P^{\mathrm{s}}_{1,0}$	0.99483	0.00023	0.00029	0.00092	0.00016	0.00031	0.00324	0.00003
$P^{\mathrm{s}}_{1,1}$	0.00002	0.99791	0.00014	0.00026	0.00006	0.00005	0.00154	0.00002
$P^{\mathrm{s}}_{2,0}$	0.00065	0.00008	0.99684	0.00003	0.00001	0.00029	0.00002	0.00208
$P^{\mathrm{s}}_{2,1}$	0.00134	0.00015	0.00009	0.99482	0.00008	0.00028	0.00000	0.00323
$P^{\mathrm{s}}_{3,0}$	0.00008	0.00023	0.00011	0.00000	0.99883	0.00004	0.00044	0.00027
$P^{\mathrm{s}}_{3,1}$	0.00011	0.00023	0.00022	0.00016	0.00016	0.99803	0.00050	0.00061

Table A.2: Values of the weights used to define each secondary preparation procedure. Each of the six secondary preparation procedures, denoted $P^{\mathrm{s}}_{t,b}$ where $t \in \{1,2,3\}, b \in \{0,1\}$ (the rows), is a probabilistic mixture of the eight primary preparation procedures, denoted $P^{\mathrm{p}}_{t',b'}$ where $t' \in \{1,2,3,4\}, b' \in \{0,1\}$ (the columns). The table presents the weights appearing in each such mixture, denoted $u^{t,b}_{t',b'}$ in Section 4.3.4. These are determined numerically by maximizing the function $C_{\mathrm{P}} = \frac{1}{6}\sum_{t=1}^{3}\sum_{b=0}^{1} u^{t,b}_{t,b}$ (the average of the weights appearing in the shaded cells), which quantifies the closeness of the secondary procedures to the primary ones, subject to the constraint of operational equivalence of the uniform mixtures of $P^{\mathrm{s}}_{t,0}$ and $P^{\mathrm{s}}_{t,1}$ for $t \in \{1,2,3\}$. The values presented are averages over 100 runs.

The scheme for finding the weights $(v^t_1, v^t_2, v^t_3, v^t_4, v^t_{\mathbf{0}}, v^t_{\mathbf{1}}, v^t_{\neg 4})$ that define the secondary measurements is analogous. Using a linear program, we find the vector of such weights that maximizes the function

$$C_{\mathrm{M}} \equiv \frac{1}{3}\sum_{t=1}^{3} v^t_t, \tag{A.20}$$

subject to the constraint that

$$\mathbf{M}^{\mathrm{s}}_{*} = \frac{1}{2}\mathbf{1}, \tag{A.21}$$

where $\mathbf{M}^{\mathrm{s}}_{*} \equiv \frac{1}{3}\sum_{t=1}^{3}\mathbf{M}^{\mathrm{s}}_{t}$. A high value of C_{M} signals that each of the three secondary measurements is close to the corresponding primary one. Table A.3 reports the weights we obtain from this optimization procedure, averaged over the 100 runs of the experiment. These weights yield $C_{\mathrm{M}} = 0.9976 \pm 0.0001$, again indicating the closeness of the secondary measurements to the primary ones.

This optimization defines the precise linear transformation of the rows of D^{p} and the linear transformation of the columns of D^{p} that serve to define the secondary preparations and measurements. By combining the operations on the rows and on the columns, we

	$[0\|M_1^{\mathrm{p}}]$	$[0\|M_2^{\mathrm{p}}]$	$[0\|M_3^{\mathrm{p}}]$	$[0\|M_4^{\mathrm{p}}]$	$[1\|M_4^{\mathrm{p}}]$	1	0
$[0\|M_1^{\mathrm{s}}]$	0.99707	0.00004	0.00015	0.00010	0.00208	0.00031	0.00025
$[0\|M_2^{\mathrm{s}}]$	0.00007	0.99727	0.00012	0.00004	0.00199	0.00028	0.00023
$[0\|M_3^{\mathrm{s}}]$	0.00004	0.00002	0.99845	0.00001	0.00117	0.00019	0.00012

Table A.3: Values of the weights used to define each secondary measurement procedure. Each of the three secondary outcome-0 measurement events, denoted $[0|M_t^{\mathrm{s}}]$ where $t \in \{1,2,3\}$ (the rows), is a probabilistic mixture of the four primary outcome-0 measurement events, denoted $[0|M_{t'}^{\mathrm{p}}]$ where $t' \in \{1,2,3,4\}$, and three processings thereof, denoted $[1|M_4^{\mathrm{p}}]$, 1, and 0 (the seven columns). The table presents the weights appearing in each such mixture. These are determined numerically by maximizing the function $C_{\mathrm{M}} = \frac{1}{3}\sum_{t=1}^{3} v_t^t$ (the average of the weights appearing in the shaded cells), which quantifies the closeness of the secondary procedures to the primary ones, subject to the constraint of operational equivalence between the uniform mixture of M_1^{s}, M_2^{s} and M_3^{s} and a fair coin flip. The values presented are averages over 100 runs.

obtain from D^{p} a 3×6 matrix, denoted D^{s}, whose entries $s_{t,b}^{t'}$ are

$$\sum_{\tau=1}^{4}\sum_{\beta=0}^{1} u_{\tau,\beta}^{t,b}\left[\sum_{\tau'=1}^{4} v_{\tau'}^{t'} p_{\tau,\beta}^{\tau'} + v_{\mathbf{0}}^{t'}0 + v_{\mathbf{1}}^{t'}1 + v_{\neg 4}^{t'}(1-p_{\tau,\beta}^{4})\right] \tag{A.22}$$

where $t', t \in \{1,2,3\}$, $b \in \{0,1\}$. This matrix describes the secondary preparations $P_{t,b}^{\mathrm{s}}$ and measurements $M_{t'}^{\mathrm{s}}$. The component $s_{t,b}^{t'}$ of this matrix describes the probability of obtaining outcome 0 in measurement $M_{t'}^{\mathrm{s}}$ on preparation $P_{t,b}^{\mathrm{s}}$, that is,

$$s_{t,b}^{t'} \equiv p(0|M_{t'}^{\mathrm{s}}, P_{t,b}^{\mathrm{s}}). \tag{A.23}$$

These probabilities are the ones that are used to calculate the value of A via Eq. (4.6).

A.3 Data analysis

A.3.1 Fitting the raw data to a generalised probabilistic theory

Note: The fitting procedure that we describe in this section could be replaced with the GPT tomography scheme detailed in Chapter 5 and Appendix B (and in fact, both methods return

the same results). However, the experiment in Chapter 4 was performed before the work in Chapter 5, and thus we used an alternate method for finding the GPT description of our that experiment. We include the description of this alternate method here for completeness.

In our experiment we perform four measurements on each of eight input states. If we define $r^{t'}_{t,b}$ as the fraction of '0' outcomes returned by measurement $M_{t'}$ on preparation $P_{t,b}$, the results can be summarized in a 4×8 matrix of raw data, D^{r}, defined as:

$$D^{\mathrm{r}} = \begin{pmatrix} r^1_{1,0} & r^1_{1,1} & \cdots & r^1_{4,0} & r^1_{4,1} \\ r^2_{1,0} & r^2_{1,1} & \cdots & r^2_{4,0} & r^2_{4,1} \\ r^3_{1,0} & r^3_{1,1} & \cdots & r^3_{4,0} & r^3_{4,1} \\ r^4_{1,0} & r^4_{1,1} & \cdots & r^4_{4,0} & r^4_{4,1} \end{pmatrix}. \tag{A.24}$$

Each row of D^{r} corresponds to a measurement, ordered from top to bottom as M_1, M_2, M_3, and M_4. Similary, the columns are labelled from left to right as $P_{1,0}$, $P_{1,1}$, $P_{2,0}$, $P_{2,1}$, $P_{3,0}$, $P_{3,1}$,$P_{4,0}$, and $P_{4,1}$.

In order to test the assumption that three independent binary-outcome measurements are tomographically complete for our system, we fit the raw data to a matrix, D^{p}, of primary data defined in Eq. (A.13). D^{p} contains the outcome probabilities of four measurements on eight states in the GPT-of-best-fit to the raw data. We fit to a GPT in which three 2-outcome measurements are tomographically complete, which we characterize with the following result.

Proposition 1 *A matrix D^p can arise from a GPT in which three two-outcome measurements are tomographically complete if and (with a measure zero set of exceptions) only if $ap^1_{t,b} + bp^2_{t,b} + cp^3_{t,b} + dp^4_{t,b} - 1 = 0$ for some real constants $\{a, b, c, d\}$.*

Proof. We begin with the "only if" part. Following [50, 51], if a set of two-outcome measurements M_A, M_B, M_C (called *fiducial* measurements) are tomographically complete for a system, then the state of the system given a preparation P can be specified by the vector

$$\mathbf{p} = \begin{pmatrix} 1 \\ p(0|M_A, P) \\ p(0|M_B, P) \\ p(0|M_C, P) \end{pmatrix} \tag{A.25}$$

(where the first entry indicates that the state is normalized). In [50, 51] it is shown that convexity then requires that the probability of outcome '0' for any measurement M is given by $\mathbf{r} \cdot \mathbf{p}$ for some vector $\mathbf{r}$. Let $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4$ correspond to outcome '0' of the measurements

M_1, M_2, M_3, M_4, and note that the measurement event that *always* occurs, regardless of the preparation (e.g. the event of obtaining either outcome '0' or '1' in any binary-outcome measurement), must be represented by $\mathbf{r}_{\mathbb{I}} = (1, 0, 0, 0)$. Since the $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4, \mathbf{r}_{\mathbb{I}}$ are a set of five four-dimensional vectors, they must be linearly dependent:

$$a'\mathbf{r}_1 + b'\mathbf{r}_2 + c'\mathbf{r}_3 + d'\mathbf{r}_4 + e'\mathbf{r}_{\mathbb{I}} = 0 \tag{A.26}$$

with $(a', b', c', d', e') \neq (0, 0, 0, 0, 0)$. The set of $\mathbf{r}$ for which e' *must* be zero are those where $\mathbf{r}_{\mathbb{I}}$ is not in the span of $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4$, which is a set of measure zero. Hence we can generically ensure $e' \neq 0$ and divide Eq. (A.26) through by $-e'$ to obtain

$$a\mathbf{r}_1 + b\mathbf{r}_2 + c\mathbf{r}_3 + d\mathbf{r}_4 - \mathbf{r}_{\mathbb{I}} = 0 \tag{A.27}$$

where $a = -a'/e'$, $b = -b'/e'$ and so on.

Finally, letting $\mathbf{p}_{t,b}$ denote the column vector of the form of Eq. (A.25) that is associated to the preparation $P_{t,b}$, and noting that by definition

$$p^{t'}_{t,b} = \mathbf{r}_{t'} \cdot \mathbf{p}_{t,b}, \tag{A.28}$$

we see that by taking the dot product of Eq. (A.27) with each $\mathbf{p}_{t,b}$, we obtain the desired constraint on D_{p}.

For the "if" part, we assume the constraint and demonstrate that there exists a triple of binary-outcome measurements, M_A, M_B, and M_C, that are tomographically complete for the GPT. To establish this, it is sufficient to take the fiducial set, M_A, M_B and M_C, to be M_1, M_2, and M_3, so that preparation $P_{t,b}$ corresponds to the vector

$$\mathbf{p}_{t,b} = \begin{pmatrix} 1 \\ p^1_{t,b} \\ p^2_{t,b} \\ p^3_{t,b} \end{pmatrix}. \tag{A.29}$$

In this case, we can recover D^{p} if M_1, M_2, and M_3 are represented by $\mathbf{r}_1 = (0, 1, 0, 0)$, $\mathbf{r}_2 = (0, 0, 1, 0)$ and $\mathbf{r}_3 = (0, 0, 0, 1)$, whilst the assumed constraint implies that $\mathbf{r}_4 = -(-1, a, b, c)/d$. ■

Geometrically, the proposition dictates that the eight columns of D^{p} lie on the 3-dimensional hyperplane defined by the constants $\{a, b, c, d\}$.

To find the GPT-of-best-fit we fit a 3-d hyperplane to the eight 4-dimensional points that make up the columns of D^{r}. We then map each column of D^{r} to its closest point on the

hyperplane, and these eight points will make up the columns of D^{p}. We use a weighted total least-squares procedure [96, 97] to perform this fit. Each element of D^{r} has an uncertainty, $\Delta r^{t'}_{t,b}$, which is estimated assuming the dominant source of error is the statistical error arising from Poissonian counting statistics. We define the *weighted distance*, $\chi_{t,b}$, between the (t,b) column of D^{r} and D^{p} as $\chi_{t,b} = \sqrt{\sum_{t'=1}^{4} \left(r^{t'}_{t,b} - p^{t'}_{t,b}\right)^2 / \left(\Delta r^{t'}_{t,b}\right)^2}$. Finding the best-fitting hyperplane can be summarized as the following minimization problem:

$$\begin{aligned} \underset{\{p^i_{t,b},a,b,c,d\}}{\text{minimize}} \quad & \chi^2 = \sum_{t=1}^{4}\sum_{b=0}^{1} \chi^2_{t,b}, \\ \text{subject to} \quad & a p^1_{t,b} + b p^2_{t,b} + c p^3_{t,b} + d p^4_{t,b} - 1 = 0 \\ & \forall\, t = 1,\ldots,4, b = 0, 1. \end{aligned} \tag{A.30}$$

The optimization problem as currently phrased is a problem in 36 variables—the 32 elements of D^{p} together with the hyperplane parameters $\{a,b,c,d\}$. We can simplify this by first solving the simpler problem of finding the weighted distance $\chi_{t,b}$ between the (t,b) column of D^{r} and the hyperplane $\{a,b,c,d\}$. This can be phrased as the following 8-variable optimization problem:

$$\begin{aligned} \underset{\{p^1_{t,b},p^2_{t,b},p^3_{t,b},p^4_{t,b}\}}{\text{minimize}} \quad & \chi^2_{t,b} = \sum_{t'=1}^{4} \frac{(r^{t'}_{t,b} - p^{t'}_{t,b})^2}{\left(\Delta r^{t'}_{t,b}\right)^2}, \\ \text{subject to} \quad & a p^1_{t,b} + b p^2_{t,b} + c p^3_{t,b} + d p^4_{t,b} - 1 = 0. \end{aligned} \tag{A.31}$$

Using the method of Lagrange multipliers [96], we define the Lagrange function $\Gamma = \chi^2_{t,b} + \gamma(a p^1_{t,b} + b p^2_{t,b} + c p^3_{t,b} + d p^4_{t,b} - 1)$, where γ denotes the Lagrange multiplier, then simultaneously solve

$$\frac{\partial \Gamma}{\partial \gamma} = 0 \tag{A.32}$$

and

$$\frac{\partial \Gamma}{\partial p^{t'}_{t,b}} = 0,\ t' = 1,\ldots,4 \tag{A.33}$$

for the variables γ, $p^1_{t,b}$, $p^2_{t,b}$, $p^3_{t,b}$, and $p^4_{t,b}$. Substituting the solutions for $p^1_{t,b}$, $p^2_{t,b}$, $p^3_{t,b}$ and $p^4_{t,b}$ into Eq. (A.31) we find

$$\chi^2_{t,b} = \frac{(a r^1_{t,b} + b r^2_{t,b} + c r^3_{t,b} + d r^4_{t,b} - 1)^2}{\left(a\Delta r^1_{t,b}\right)^2 + \left(b\Delta r^2_{t,b}\right)^2 + \left(c\Delta r^3_{t,b}\right)^2 + \left(d\Delta r^4_{t,b}\right)^2}, \tag{A.34}$$

which now only contains the variables a, b, c, and d.

The hyperplane-finding problem can now be stated as the following four-variable optimization problem:

$$\underset{\{a,b,c,d\}}{\text{minimize}} \quad \chi^2 = \sum_{t=1}^{4} \sum_{b=0}^{1} \chi^2_{t,b} \tag{A.35}$$

which we solve numerically.

The χ^2 parameter returned by the fitting procedure is a measure of the goodness-of-fit of the hyperplane to the data. Since we are fitting eight data points to a hyperplane defined by four fitting parameters $\{a, b, c, d\}$, we expect the χ^2 parameter to be drawn from a χ^2 distribution with four degrees of freedom [97], which has a mean of 4. As stated in Chapter 4, we ran our experiment 100 times and obtained 100 independent χ^2 parameters; these have a mean of 3.9 ± 0.3. In addition we performed a more stringent test of the fit of the model to the data by summing the counts from all 100 experimental runs before performing a single fit. This fit returns a χ^2 of 4.33, which has a p-value of 36%. The outcomes of these tests are consistent with our assumption that the raw data can be explained by a GPT in which three 2-outcome measurements are tomographically complete and which also exhibits Poissonian counting statistics. Had the fitting procedure returned χ^2 values that were much higher, this would have indicated that the theoretical description of the preparation and measurement procedures required more than three degrees of freedom. On the other hand, had the fitting returned an average χ^2 much lower than 4, this would have indicated that we had overestimated the amount of uncertainty in our data.

After finding the hyperplane-of-best-fit $\{a, b, c, d\}$, we find the points on the hyperplane that are closest to each column of D^{r}. This is done by numerically solving for $p^1_{t,b}$, $p^2_{t,b}$, $p^3_{t,b}$, and $p^4_{t,b}$ in (A.31) for each value of (t, b). The point on the hyperplane closest to the (t, b) column of D^{r} becomes the (t, b) column of D^{p}. The matrix D^{p} is then used to find the secondary preparations and measurements.

A.3.2 Why is fitting to a GPT necessary?

It is clear that one needs to assume that the measurements one has performed form a tomographically complete set, otherwise statistical equivalence relative to those measurements does not imply statistical equivalence relative to all measurements. (Recall that the assumption of preparation noncontextuality only has nontrivial consequences when two preparations are statistically equivalent for all measurements.)

The minimal assumption for our experiment would therefore be that the four measurements we perform are tomographically complete. But our physical understanding of the experiment leads us to a stronger assumption, that three measurements are tomographically complete. Here we clarify why, given this latter assumption, it is necessary to carry out the step of fitting to an appropriate GPT.

It is again easier to begin by considering the case that our experiment is described by quantum theory. Let $(q^1_{t,b}, q^2_{t,b}, q^3_{t,b}, q^4_{t,b})$ denote the probability of obtaining outcome '0' in measurements M_1, M_2, M_3, M_4 on preparation $P_{t,b}$, according to quantum theory, namely $q^i_{t,b} = \mathrm{Tr}(E_i \rho_{t,b})$, where E_i is the POVM element corresponding the the 0 outcome of measurement M_i and $\rho_{t,b}$ is the density operator for $P_{t,b}$.

Let us represent $\rho_{t,b} = \frac{1}{2}(\mathbb{I} + \boldsymbol{\sigma} \cdot \boldsymbol{u}_{t,b})$ by a Bloch vector $\boldsymbol{u}_{t,b}$ and the elements $E_i = v^0_i \mathbb{I} + \boldsymbol{\sigma} \cdot \boldsymbol{v}_i$ by a "Bloch four-vector" $(v^0_i, \boldsymbol{v}_i)$. Then $q^i_{t,b} = v^0_i + \boldsymbol{u}_{t,b} \cdot \boldsymbol{v}_i$. Since the $\boldsymbol{v}_i$ lie in a unit sphere, the $(q^1_{t,b}, q^2_{t,b}, q^3_{t,b}, q^4_{t,b})$ lie in the image of the sphere under the affine transformation $\boldsymbol{u} \mapsto (v^0_1, v^0_2, v^0_3, v^0_4) + (\boldsymbol{v}_1 \cdot \boldsymbol{u}, \boldsymbol{v}_2 \cdot \boldsymbol{u}, \boldsymbol{v}_3 \cdot \boldsymbol{u}, \boldsymbol{v}_4 \cdot \boldsymbol{u})$, i.e. some ellipsoid, a three-dimensional shape in a four-dimensional space.

However, the relative frequencies we observe will fluctuate from $q^i_{t,b}$ in all four dimensions. Fluctuations in the three dimensions spanned by the "Bloch ellipsoid" can be accommodated by using secondary preparations as described above. But fluctuations in the fourth direction are, according to quantum theory, always statistical and never systematic, and by the same token we cannot deliberately produce supplementary preparations that have any bias in this fourth direction. Therefore, we need to deal with these fluctuations in a different way. If one was assuming quantum theory, one would simply fit relative frequencies to the closest points $q^{t'}_{t,b}$ in the Bloch ellipsoid, just as one usually fits to the closest valid density operator.

Since we do not assume quantum theory, we do not assume that the states lie in an ellipsoid. However, we still make the assumption that three two-outcome measurements are tomographically complete. Hence, by Proposition 1, the long-run probabilities lie in a three-dimensional subspace of a four-dimensional space, and so there are no supplementary preparations that can deal with fluctuations of relative frequencies in the fourth dimension. Instead of fitting to the "Bloch ellipsoid", we fit to a suitable GPT.

A.3.3 Analysis of statistical errors

Because the relative frequencies derived from the raw data constitute a finite sample of the true probabilities (i.e. the long-run relative frequencies), the GPT states and effects

that yield the best fit to the raw data are *estimates* of the GPT states and effects that characterize the primary preparations and measurements.

It is these estimates that we input into the linear program that identifies the weights with which the primary procedures must be mixed to yield secondary procedures. As such, our linear program outputs estimates of the true weights, and therefore when we use these weights in mixtures of our estimates of the primary GPT states and effects, we obtain estimates of the secondary GPT states and effects. In turn, these estimates are input into the expression for A and yield an estimate of the value of A for the secondary preparations and measurements.

To determine the statistical error on our estimate of A, we must quantify the statistical error on our estimates of the GPT states for the primary preparations and on our estimates of the GPT effects for the primary measurements. We do so by taking our experimental data in 100 distinct runs, each of which yields one such estimate. For each of these, we follow the algorithm for computing the value of A. In this way, we obtain 100 samples of the value of A for the secondary procedures, and these are used to determine the statistical error on our estimate for A.

Note that a different approach would be to presume some statistical noise model for our experiment, then input the observed relative frequencies (averaged over the entire experiment) into a program that adds noise using standard Monte Carlo techniques. Though one could generate a greater number of samples of A in this way, such an approach would be worse than the one we have adopted because the error analysis would be only as reliable as one's assumptions regarding the nature of the noise.

Given that the quantity A we obtain is 2300 σ above the noncontextual bound, we can conclude that there is a very low likelihood that a noncontextual model would provide a better fit to the true probabilities than the GPT that best fit our finite sample would. This is the sense in which our experiment rules out a noncontextual model with high confidence.

It should be noted that this sort of analysis of statistical errors is no different from that which has historically been used for experimental tests of Bell inequalities. The Bell quantity (the expression that is bounded in a Bell inequality) is defined in terms of the true probabilities. Any Bell experiment, however, only gathers a finite sample of these true probabilities. From this sample, one estimates the true probabilities and in turn the value of the Bell quantity. We treat the quantity A appearing in our noncontextuality inequality in a precisely analogous manner. The definition of A in terms of the true probabilities is admittedly more complicated than for a Bell quantity: we define secondary procedures based on an optimization problem that takes as input the true probabilities for the primary procedures, and use the true probabilites for the secondary procedures to define A. But

this complication does not change the fact that A is ultimately just a function of the true probabilities for the primary preparations and measurements, albeit a function that incorporates a particular linear optimization problem in its definition.

Recently, more sophisticated statistical techniques have been applied to the analysis of tests of Bell inequalities [159, 160, 161, 162, 163, 164]. Specifically, one computes an upper bound on the probability that a locally causal model could reproduce the Bell quantity observed in the experiment. This techniques has been applied to the analysis of the recent loophole-free violations of Bell inequalities [153, 155, 154]. It would be worthwhile to make a similar analysis of our experiment, by computing an upper bound on the probability that a noncontextual model could reproduce the value of A we observe. Such an analysis is outside the scope of the present work, but an interesting problem for future work in this area.

Appendix B

Appendices for Chapter 5

Notes and acknowledgements

This chapter contains the appendices of work that has appeared on the arXiv as [1]:

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," *arXiv:1710.05948*, 2017.

Author contributions

M. D. Mazurek and **R. W. Spekkens** wrote the first draft of the appendices.

All authors contributed to the final draft.

B.1 Methods

B.1.1 Photon source

The 20 mm long PPKTP crystal is pumped with 0.29 mW of continuous wave laser light at 404.7 nm, producing pairs of 809.4 nm photons with orthogonal polarizations. We detect approximately 22% of the herald photons produced, and approximately 9% of the signal

photons produced. In order to characterize the single-photon nature of the source we performed a $g^2(0)$ measurement [11] and found $g^2(0) = 0.00184 \pm 0.00003$. This low $g^2(0)$ measurement implies that the ratio of double pairs to single pairs produced by the source is $\sim 1 : 2000$. We found that if we increased the pump power then a rank 4 model no longer fit the data well. This is because the two-photon state space has a higher dimension than the one-photon state space. The avalanche photodiode single photon detectors we use respond nonlinearly to the number of incoming photons [16]; this makes our measurements sensitive to the multi-pair component of the downconverted light and ultimately limits the maximum power we can set for the pump laser.

B.1.2 Measurements

After a photon exits the measurement PBS, the probability that it will be detected depends on which port it exited the PBS from. This is because the efficiencies of the two paths from the measurement PBS to the detector are not exactly equal, and also because the detectors themselves do not have the same efficiency. To average out the two different efficiencies we perform each measurement in two stages. First, we rotate the measurement quarter and half waveplates to the angles for which photons with some polarization $|\psi\rangle$ will be *transmitted* by the measurement PBS, and in each output port record the number of photons detected in coincidence with the herald, for four seconds. We label detections in the transmitted port with '0' and detections in the reflected port with '1'. Second, we rotate the measurement waveplates so $|\psi\rangle$ will be *reflected* at the measurement PBS, and swap the labels on the measurement outcomes such that the reflected port corresponds to outcome '0' and the transmitted port to '1'. We record the number of coincidences with the herald in each output port for another four seconds, and then sum the total number of '0' detections, and also the number '1' detections over the total eight-second measurement time.

B.2 Choice of preparation and measurement settings

We choose the preparation and measurement settings in our experiment with the aim of characterizing the largest volume of the state and measurement effect spaces as possible. The state and effect spaces in any GPT are convex, and thus fully characterizing the boundaries of these spaces fully determines the full spaces. Thus our aim is to find preparation and measurement settings that map out the boundaries of the state and effect spaces as best we can, given the finite number of settings we are able to perform.

We use quantum theory to inform our choice of settings. We expect the GPT describing our experiment to be equal to (or very closely approximated by) the GPT for a qubit. The surface of the Bloch sphere (i.e. the space of pure qubit states) determines the qubit state space, and preparing a set states of states that are evenly distributed around the surface of the Bloch sphere should do a good job at characterizing the GPT state space describing our experiment. The qubit effect space is characterized by the surface of the sphere representing projective measurement effects, plus the unit effect, $\mathbb{I}$, and its complement, the zero effect. Thus, we aim to perform a set of measurements whose effects are evenly distributed on the outside of the sphere of projective effects.

To choose the preparation settings we first find a set of pure quantum states labelled with $|\psi_i\rangle$ that are approximately evenly distributed around the surface of the Bloch sphere. We then find the quarter and half waveplate angles necessary to create each of those states, and each pair of quarter and half waveplate angles is one preparation setting. The space of projective effects is also determined by the Bloch sphere, since every projective effect $|\psi_i\rangle\langle\psi_i|$ can be associated with the state to which it responds deterministically, $|\psi_i\rangle$. The measurement settings are the waveplate angles that implement the projective measurements $\{|\psi_i\rangle\langle\psi_i|, \mathbb{I} - |\psi_i\rangle\langle\psi_i|\}$.

We use a method due to Rakhmanov, Saff, and Zhou [165] to find the set of approximately uniformly distributed points on the surface of the Bloch sphere. The points lie on a spiral that begins at the south pole of the sphere, and winds up around the sphere and ends at the north pole. The quantum states corresponding to each of the 100 preparation settings in the first experiment are shown in Fig. B.1(a), and the 1000 states corresponding to each preparation setting in the second experiment are displayed in Fig. B.1(b).

In the second experiment, we also implement a set of six fiducial preparations which we use to characterize each of the 1000 effects in Fig. B.1(b), and a set of six fiducial measurements which we use to characterize each of the 1000 states in Fig. B.1(b). The fiducial preparation and measurement sets are shown in Fig. B.2.

B.3 Finding the rank-k matrix $\tilde{D}$ that best fits the frequency matrix F

In this section we explain the algorithm we use to find a low-rank matrix that best fits the matrix of raw frequency data.

For an $m \times n$ matrix of frequency data, F, we define the rank-k matrix of best fit, $\tilde{D}$,

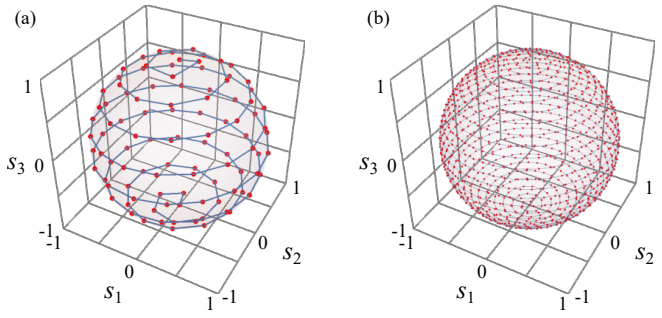

Figure B.1: Quantum description of the target states created and measurements performed in our experiment. An evenly distributed set of points lying on a spiral was used to choose the settings for (a) the 100 preparations and measurements characterized in the first experiment and measurements and (b) the 1000 nonfiducial preparations and measurements characterized in the second experiment. Each red dot corresponds to a quantum state $|\psi_i\rangle$, and the waveplate angles (i.e. preparation settings) were chosen in order to prepare those states. Each red dot also defines an effect $|\psi_i\rangle\langle\psi_i|$ which is part of the projective measurement $\{|\psi_i\rangle\langle\psi_i|, \mathbb{I} - |\psi_i\rangle\langle\psi_i|\}$.

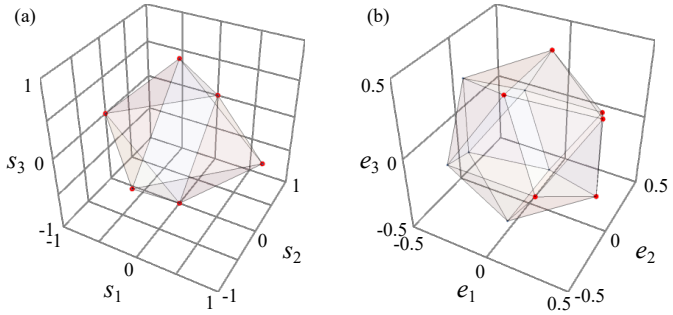

Figure B.2: Quantum description of the fiducial states and measurement effects performed in the second experiment. (a) Red dots represent the six fiducial states used to characterize the 1000 measurements in Fig. B.1(b). These correspond to the +1 and -1 eigenstates of the three Pauli operators σ_x, σ_y, and σ_z. (b) Red dots represent the six fiducial measurement effects used to characterize each of the states in Fig. B.1(b). These effects lie on six of the twelve vertices of an icosahedron, and they correspond to the outcome-'0' effect of a projective measurement. Each outcome-'0' effect has a corresponding outcome-'1' effect; each outcome-'1' effect is represented by one of the other six vertices on the icosahedron.

as the one that minimizes the weighted χ^2 value:

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\frac{F_{ij} - \tilde{D}_{ij}}{\Delta F_{ij}} \right)^2, \tag{B.1}$$

where the weights ΔF_{ij} are the uncertainties in the measured frequencies, which are calculated assuming Poissonian error in the counts (in cases where we did not collect data for the preparation-measurement pair corresponding to entry F_{ij}, we set $\Delta F_{ij} = \infty$). Since $\tilde{D}$ represents an estimate of the true probabilities underlying the noisy frequency data, we need to ensure that $\tilde{D}$ only contains entries between 0 and 1. Hence the matrix of best fit is the one which solves the following minimization problem:

$$\begin{aligned} \underset{\tilde{D} \in M_{mn}}{\text{minimize}} \quad & \chi^2, \\ \text{subject to} \quad & \operatorname{rank}(\tilde{D}) \leq k \\ & 0 \leq \tilde{D}_{ij} \leq 1 \quad \forall\, i, j, \end{aligned} \tag{B.2}$$

where M_{mn} is the space of all $m \times n$ real matrices. The entries in the column of ones (representing the unit measurement effect) that we include in F are *exact*, meaning that they have an uncertainty of 0. As $\tilde{D}$ is defined as the matrix that minimizes χ^2, this enforces that the entries in the same column of $\tilde{D}$ will also remain exactly 1. Otherwise, χ^2 would be undefined.

To enforce the rank constraint, we use the parameterization $\tilde{D} = \tilde{S}\tilde{E}$, where $\tilde{S}$ has size $m \times k$ and $\tilde{E}$ is $k \times n$. This minimization problem as stated is NP-hard [138], and cannot be solved analytically. However, if either $\tilde{S}$ or $\tilde{E}$ remains fixed, optimizing the other variable is a convex problem which can be solved with a quadratic program. We minimize χ^2 by performing a series of alternating optimizations over $\tilde{S}$ and $\tilde{E}$ [139].

Each iteration begins with an estimate for $\tilde{E}$, and we find the $\tilde{S} = \tilde{S}'$ which minimizes the χ^2. Next, we fix $\tilde{S} = \tilde{S}'$ and find the optimal $\tilde{E} = \tilde{E}'$. This is the end of one iteration, and $\tilde{E}'$ becomes the fixed value of $\tilde{E}$ for the beginning of the next iteration. The algorithm runs until a specific convergence threshold is met (i.e. if $\Delta\chi^2 < 10^{-6}$ between successive iterations), or until a maximum number of iterations (we choose 5000) is reached.

We will now show that optimization over $\tilde{S}$ or $\tilde{E}$ is convex (given that the other variable is fixed). For what follows, we will make use of the vec$(\cdot)$ operator, which takes a matrix and reorganises its entries into a column vector with the same number of entries as the original matrix. For example, given an $m \times n$ matrix A, vec(A) is a vector of length mn, and the first m entries of vec(A) are equal to the first column of A, entries $m+1$ through

$2m$ are equal to the second column of A, and so on. We also define a diagonal $mn \times mn$ matrix of weights, W, to encode the uncertainties $(1/\Delta F_{ij})^2$. These values appear along the diagonal of W, and they are appropriately ordered such that we can rewrite χ^2 in the more convenient form:

$$\begin{aligned}\chi^2 &= \operatorname{vec}(F - \tilde{S}\tilde{E})^T W \operatorname{vec}(F - \tilde{S}\tilde{E}) && \text{(B.3)}\\ &= \operatorname{vec}(\tilde{S}\tilde{E})^T W \operatorname{vec}(\tilde{S}\tilde{E}) - 2\operatorname{vec}(\tilde{S}\tilde{E})^T W \operatorname{vec}(F) \\ &\quad + \operatorname{vec}(F)^T W \operatorname{vec}(F), && \text{(B.4)}\end{aligned}$$

where we have also made the substitution $\tilde{D} = \tilde{S}\tilde{E}$.

Defining I_m as the $m \times m$ identity matrix, we can use the identity $\operatorname{vec}(\tilde{S}\tilde{E}) = (\tilde{E} \otimes I_m)\operatorname{vec}(\tilde{S})$ to write:

$$\begin{aligned}\chi^2 = &\operatorname{vec}\left(\tilde{S}\right)^T (\tilde{E} \otimes I_m)^T W (\tilde{E} \otimes I_m) \operatorname{vec}\left(\tilde{S}\right) \\ &- 2\operatorname{vec}\left(\tilde{S}\right)^T (\tilde{E} \otimes I_m)^T W \operatorname{vec}(F) \\ &+ \operatorname{vec}(F)^T W \operatorname{vec}(F), \end{aligned} \qquad \text{(B.5)}$$

and we now see that the minimization over P can be written as:

$$\begin{aligned}\underset{\tilde{S}\in M_{mk}}{\text{minimize}} \quad & \operatorname{vec}\left(\tilde{S}\right)^T (\tilde{E} \otimes I_m)^T W (\tilde{E} \otimes I_m) \operatorname{vec}\left(\tilde{S}\right) \\ & \qquad - 2\operatorname{vec}\left(\tilde{S}\right)^T (\tilde{E} \otimes I_m)^T W \operatorname{vec}(F) \\ \text{subject to} \quad & 0 \leq (\tilde{S}\tilde{E})_{ij} \leq 1 \quad \forall\, i, j. \end{aligned} \qquad \text{(B.6)}$$

We have ignored the third term of Eq. (B.4) as it is a constant, and depends neither on $\tilde{S}$ nor $\tilde{E}$. Since W is a diagonal matrix consisting of only positive elements, $(\tilde{E} \otimes I_m)^T W (\tilde{E} \otimes I_m)$ is positive-definite. This means that (B.6) is a convex quadratic program [166] which can be solved in polynomial time.

The optimization over $\tilde{E}$ takes a similar form, which can be found by applying the identity $\operatorname{vec}(\tilde{S}\tilde{E}) = (I_n \otimes \tilde{S})\operatorname{vec}(\tilde{E})$ to Eq. (B.4):

$$\begin{aligned}\underset{\tilde{E}\in M_{kn}}{\text{minimize}} \quad & \operatorname{vec}(\tilde{E})^T (I_n \otimes S)^T W (I_n \otimes \tilde{S}) \operatorname{vec}(\tilde{E}) \\ & \qquad - 2\operatorname{vec}(\tilde{E})^T (I_n \otimes \tilde{S})^T W \operatorname{vec}(F) \\ \text{subject to} \quad & 0 \leq (\tilde{S}\tilde{E})_{ij} \leq 1 \quad \forall\, i, j. \end{aligned} \qquad \text{(B.7)}$$

B.4 Decomposition of the fitted matrix of probabilities

As discussed in Section 5.4.5 in the main paper, we find a decomposition $\tilde{D}^{\text{realized}} = \tilde{S}^{\text{realized}}\tilde{E}^{\text{realized}}$ in order to characterize the estimates of the spaces realized by the experiment, $\tilde{\mathcal{S}}_{\text{realized}}$ and $\tilde{\mathcal{E}}_{\text{realized}}$. Here, $\tilde{D}^{\text{realized}}$ has size $m \times n$, $\tilde{S}^{\text{realized}}$ is $m \times k$ and $\tilde{E}^{\text{realized}}$ is $k \times n$. In this appendix we describe the method we use to perform the above decomposition.

We choose the decomposition to ensure that the first column of $\tilde{S}^{\text{realized}}$ is a column of ones, which allows us to represent $\tilde{\mathcal{S}}_{\text{realized}}$ in $k-1$ dimensions. (In our experiment we found $k = 4$, but we will use the symbol k in this appendix for generality.) We achieve this by ensuring that the leftmost column in $\tilde{D}^{\text{realized}}$ is a column of ones representing the unit measurement, such that $\tilde{D}^{\text{realized}}$ takes the form:

$$\tilde{D}^{\text{realized}} = \begin{pmatrix} 1 & p(0|P_1, M_2) & \cdots & p(0|P_1, M_n) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & p(0|P_m, M_2) & \cdots & p(0|P_m, M_n) \end{pmatrix}. \tag{B.8}$$

We then proceed to perform the QR decomposition [167] $\tilde{D}^{\text{realized}} = QR$, where R is an $m \times n$ upper-right triangular matrix and Q an $m \times m$ unitary matrix. Because $\tilde{D}^{\text{realized}}$ has the form of Eq. (B.8), each entry in the first column of Q will be equal to some constant c. We define $Q' = Q/c$ and $R' = cR$, which ensures that the first column of Q' is a column of ones.

Next, we partition Q' and R' as $Q' = \begin{pmatrix} Q_0 & Q_1 \end{pmatrix}$ and $R' = \begin{pmatrix} R_0 \\ R_1 \end{pmatrix}$, where Q_0 is the first column of Q', Q_1 is all remaining columns of Q', R_0 is the first row of R', and R_1 is all remaining rows of R'. We take the singular value decomposition $Q_1 R_1 = U\Sigma V^T$. $Q_1 R_1$ is rank-$(k-1)$, and thus only has $(k-1)$ nonzero singular values. Hence we can partition U, Σ, and V as $U = \begin{pmatrix} U_{k-1} & U_{(k-1)\perp} \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{k-1} & 0 \\ 0 & 0 \end{pmatrix}$, and $V = \begin{pmatrix} V_{k-1} & V_{(k-1)\perp} \end{pmatrix}$. Here Σ_{k-1} is the upper-left $(k-1) \times (k-1)$ corner of Σ, and U_{k-1} and V_{k-1} are the leftmost $(k-1)$ columns of U and V, respectively. Finally, we define $\tilde{S}^{\text{realized}}$ and $\tilde{E}^{\text{realized}}$ as $\tilde{S}^{\text{realized}} = \begin{pmatrix} Q_0 & U_{k-1}\sqrt{\Sigma_{k-1}} \end{pmatrix}$ and $\tilde{E}^{\text{realized}} = \begin{pmatrix} R_0 \\ \sqrt{\Sigma_{k-1}} V_{k-1}^T \end{pmatrix}$.

The procedure described above ensures that $\tilde{S}^{\text{realized}}$ and $\tilde{E}^{\text{realized}}$ take the forms:

$$\tilde{S}^{\text{realized}} = \begin{pmatrix} 1 & s_1^{(1)} & \cdots & s_{k-1}^{(1)} \\ 1 & s_1^{(2)} & \cdots & s_{k-1}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & s_1^{(m)} & \cdots & s_{k-1}^{(m)} \end{pmatrix}, \tag{B.9}$$

and

$$\tilde{E}^{\text{realized}} = \begin{pmatrix} 1 & e_0^{(2,0)} & \cdots & e_0^{(n,0)} \\ 0 & e_1^{(2,0)} & \cdots & e_1^{(n,0)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & e_{k-1}^{(2,0)} & \cdots & e_{k-1}^{(n,0)} \end{pmatrix}, \tag{B.10}$$

where $s_t^{(u)}$ is the t-th element of the GPT state vector representing the u-th preparation, and $e_t^{(t,v)}$ is the t-th element of the GPT effect vector representing the v-th outcome of the u-th measurement.

B.4.1 Convex closure under convex mixtures and classical post-processing of $\tilde{E}^{\text{realized}}$

As discussed in Section 5.4.5, $\tilde{\mathcal{E}}_{\text{realized}}$ is obtained by considering the convex closure under convex mixtures and classical post-processing of $\tilde{E}^{\text{realized}}$. We only perform two-outcome measurements in our experiment, and thus the full set of effects in $\tilde{\mathcal{E}}_{\text{realized}}$ is the convex hull of the outcome-0 effects of all measurement procedures implemented in the experiment (i.e. the matrix $\tilde{E}^{\text{realized}}$) *and* of all the outcome-1 effects of all the implemented measurements (i.e the matrix 1-$\tilde{E}^{\text{realized}}$).

If we chose to, we could simply take the $\tilde{E}^{\text{realized}}$ returned by the decomposition of $\tilde{D}^{\text{realized}}$ that we described above, and define the larger matrix $\begin{pmatrix} \tilde{E}^{\text{realized}} & 1 - \tilde{E}^{\text{realized}} \end{pmatrix}$, and the convex hull of the vectors in this larger matrix would define our estimate, $\tilde{\mathcal{E}}_{\text{realized}}$, of the space of GPT effects realized in the experiment.

However, in an attempt to treat the outcome-0 and outcome-1 effect vectors on equal footing, we instead define the larger matrix $\tilde{D}^{\text{R}} = \begin{pmatrix} \tilde{D}^{\text{realized}} & 1 - \tilde{D}^{\text{realized}} \end{pmatrix}$. We then find a decomposition $\tilde{D}^{\text{R}} = \tilde{S}^{\text{realized}} \tilde{E}^{\text{R}}$ using the method described above. This ensures that $\tilde{E}^{\text{R}}$

has the form:

$$\tilde{E}^{\mathrm{R}} = \begin{pmatrix} 1 & e_0^{(2,0)} & \cdots & e_0^{(n,0)} & 0 & e_0^{(2,1)} & \cdots & e_0^{(n,1)} \\ 0 & e_1^{(2,0)} & \cdots & e_1^{(n,0)} & 0 & e_1^{(2,1)} & \cdots & e_1^{(n,1)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \cdots \\ 0 & e_{k-1}^{(2,0)} & \cdots & e_{k-1}^{(n,0)} & 0 & e_{k-1}^{(2,1)} & \cdots & e_{k-1}^{(n,1)} \end{pmatrix}. \tag{B.11}$$

B.5 Calculation of dual spaces

The spaces $\tilde{\mathcal{S}}_{\text{consistent}}$ and $\tilde{\mathcal{E}}_{\text{consistent}}$ are the duals of the realized spaces $\tilde{\mathcal{E}}_{\text{realized}}$ and $\tilde{\mathcal{S}}_{\text{realized}}$, respectively. Here we will discuss how we calculate the consistent spaces from the realized ones.

We start with the calculation of $\tilde{\mathcal{S}}_{\text{consistent}}$. By definition, $\tilde{\mathcal{S}}_{\text{consistent}}$ is the intersection of the geometric dual of $\tilde{\mathcal{E}}$ set of all $\mathbf{s} \in \mathbb{R}^k$ such that $\forall \mathbf{e} \in \tilde{\mathcal{E}}_{\text{realized}} : 0 \leq \mathbf{s} \cdot \mathbf{e} \leq 1$ and $\mathbf{s} \cdot \mathbf{u} = 1$. (In our experiment, we determined that $k = 4$, but we will use the symbol k here for generality.) This definition (called an *inequality representation*) completely specifies $\tilde{\mathcal{S}}_{\text{consistent}}$. However, in order to perform transformations on the space or calculate its volume it can be useful to have its *vertex description* as well, which is a list of vertices that completely specify the space's convex hull. Finding a convex polytope's vertex representation given its inequality representation is called the *vertex enumeration problem* [168].

To find the vertex representation of $\tilde{\mathcal{S}}_{\text{consistent}}$ we first simplify its inequality representation. Since $\tilde{\mathcal{E}}_{\text{realized}}$ is a convex polytope, we don't need to consider every $\mathbf{e}$ in $\tilde{\mathcal{E}}_{\text{realized}}$, but only the vertices of $\tilde{\mathcal{E}}_{\text{realized}}$. If we define the vertices of $\tilde{\mathcal{E}}_{\text{realized}}$ as vertices $\left(\tilde{\mathcal{E}}_{\text{realized}}\right)$, then we can replace the $\forall \mathbf{e} \in \tilde{\mathcal{E}}_{\text{realized}}$ in the definition of $\tilde{\mathcal{S}}_{\text{consistent}}$ with $\forall \mathbf{e} \in \text{vertices}\left(\tilde{\mathcal{E}}_{\text{realized}}\right)$. Calculation of vertices $\left(\tilde{\mathcal{E}}_{\text{realized}}\right)$ is performed with the pyparma [169] package in Python 2.7.6. The calculation of the vertex description of $\tilde{\mathcal{S}}_{\text{consistent}}$ is performed with an algorithm provided by Avis and Fukuda [168]. We use functions in pyparma [169] which call the cdd library [170] to find the vertex description of $\tilde{\mathcal{S}}_{\text{consistent}}$.

Finding the vertex description of $\tilde{\mathcal{E}}_{\text{consistent}}$ from $\tilde{\mathcal{S}}_{\text{realized}}$ is done in an analogous way. $\tilde{\mathcal{E}}_{\text{consistent}}$ is defined as the geometric dual of the space that is the subnormalization of $\tilde{\mathcal{S}}_{\text{realized}}$, $\{w\mathbf{s} : \mathbf{s} \in \tilde{\mathcal{S}}_{\text{realized}}, w \in [0,1]\}$. The subnormalization of $\tilde{\mathcal{S}}_{\text{realized}}$ is also the convex hull of the union of the GPT state vectors that make up the rows of $\tilde{S}^{\text{realized}}$ and the GPT state vector with $s_0 = \cdots = s_{k-1} = 0$ that represents the state with normalization zero.

B.6 Relating CHSH inequality violations to POM noncontextuality inequality violations

We here provide a rigorous proof of the fact that the optimal violation of the CHSH inequality when Bob's system is described by a qubit-like state and effect space, is simply the optimal violation of the POM noncontextuality inequality.

The preparation P^{AB} followed by a measurement procedure M_x^A on Alice's system that yields outcome a defines an effective preparation procedure on Bob's system, which we denote by $P^B_{a|x}$. By this definition, the probability of observing outcome b for a measurement of M_y^B on Bob's system given a preparation $P^B_{a|x}$ is precisely equal to the probability of observing an outcome b for a measurement of M_y^B on Bob's system given an outcome a for a measurement M_x^A on Alice's system and a preparation P^{AB}. Concretely,

$$p(b|M_y^B, P^B_{a|x}) = p(b|a, M_x^A, M_y^B, P^{AB}) \tag{B.12}$$

where by the definition of conditionals,

$$p(b|a, M_x^A, M_y^B, P^{AB}) = \frac{p(a, b|M_x^A, M_y^B, P^{AB})}{p(a|M_x^A, M_y^B, P^{AB})}. \tag{B.13}$$

Given the absence of superluminal signalling from Bob to Alice, $p(a|M_x^A, M_y^B, P^{AB}) = p(a|M_x^A, P^{AB})$, and we infer that

$$p(b|a, M_x^A, M_y^B, P^{AB}) = \frac{p(a, b|M_x^A, M_y^B, P^{AB})}{p(a|M_x^A, P^{AB})}. \tag{B.14}$$

We now make use of the fact that to obtain the quantum maximum in qubit-like state and effect spaces, the marginal $p(a|M_x^A, P^{AB})$ must be uniform [171].

In the case where the marginal $p(a|M_x^A, P^{AB})$ is uniform, we have

$$\frac{1}{2} p(b|M_y^B, P^B_{a|x}) = p(a, b|M_x^A, M_y^B, P^{AB}), \tag{B.15}$$

so that

$$\mathcal{B} \equiv \frac{1}{8} \sum_{a,b,x,y} \delta_{a\oplus b,xy} p(b|M_y^B, P^B_{a|x}). \tag{B.16}$$

Finally, if we define

$$z_0 = a, \quad z_1 = a \oplus x, \tag{B.17}$$

so that the condition $a \oplus b = xy$ becomes $b = z_y$, and if we write $P^B_{a|x}$ as $P^B_{z_0 z_1}$, then we find

$$\mathcal{B} \equiv \frac{1}{8} \sum_{b,y,z_0,z_1} \delta_{b,z_y} p(b|M^B_y, P^B_{z_0 z_1}), \tag{B.18}$$

which is precisely the form of the left-hand side of the POM noncontextuality inequality, that is, the quantity C, defined in Eq. (5.4).

The no-signalling condition also has a natural expression in terms of the effective preparations on Bob's side that are steered to, namely, that the weighted average over a of the GPT states associated to the $P^B_{a|x}$ is independent of x (see Eq. (5.26)). Given that $p(a|M^A_x, P^{AB}) = \frac{1}{2}$, the constraint on the GPT states is

$$\frac{1}{2}\mathbf{s}_{P^B_{0|0}} + \frac{1}{2}\mathbf{s}_{P^B_{1|0}} = \frac{1}{2}\mathbf{s}_{P^B_{0|1}} + \frac{1}{2}\mathbf{s}_{P^B_{1|1}}. \tag{B.19}$$

If we write $P^B_{a|x}$ as $P^B_{z_0 z_1}$ under the mapping of (B.17), we recover precisely the parity-obliviousness condition of Eq. (5.5).

Therefore, the existence in a theory of a CHSH experiment implemented at space-like separation that achieves $\mathcal{B} = \alpha$ for some α implies the existence in the theory of a prepare-and-measure experiment where the preparations satisfy the the parity-obliviousness condition of Eq. (5.5) and that achieves $\mathcal{C} = \alpha$.

Appendix C

Sources of error in prepare-and-measure experiments that use polarization-encoded single photons

When running the polarization-encoded single photon prepare-and-measure experiments, I ran into some different sources of error that added additional "false" degrees of freedom to the data. It was important that the data was without these extra degrees of freedom, otherwise the GPT tomography procedure would have inferred the "wrong" dimension of GPT required to explain the data. Here I've listed a few sources of errors that gave me problems, and the steps I suggest one takes to correct them.

C.1 Waveplate-angle spatial mode coupling

I found that the preparation waveplates deflected the beam enough to affect the coupling efficiencies into the fibres leading to the detectors. This can be solved by coupling the light into a single mode fibre between the preparation and measurement stage of the experiment. This will prevent preparation-waveplate-angle-dependent deflections of the beam from affecting the spatial mode of the light that enters the measurement stage of the experiment. The fibre should be securely fixed to the table (using tape is fine) to ensure it stays in place over the course of the experiment (otherwise it will introduce additional polarization rotations to the light as it moves).

Paddle polarization controllers (bat ears) have a tendency to "fall over" during the course of the data acquisition, which introduces a time-dependent polarization rotation between the preparation and measurement stage of the experiment. It is much better to tape all fibres directly to the optical table, and use three fixed waveplates to correct fibre-induced polarization rotations.

C.2 Source drift

The photon source will drift over the course of the experiment, and thus so will the photon-pair production rate. This drift can be accounted for by counting photons in both output ports of the polarizing beamsplitter in the measurement stage.

The two detectors most likely have different efficiencies, and thus each measurement should be performed in two stages a and b. For example, if the measurement waveplates for stage a are aligned to take polarization state $|\psi\rangle$ to $|H\rangle$, then the waveplates for stage b should be aligned to take state $|\psi\rangle$ to $|V\rangle$. Then, the measured probability of obtaining outcome '0' can be calculated as

$$p(0) = \frac{n_r^a + n_t^b}{n_r^a + n_t^a + n_r^b + n_t^b}, \tag{C.1}$$

where n_r^a is the number of photons counted in the reflected port for measurement state a, and the other numbers of counts are defined similarly.

It is possible that the meaurement waveplates slightly deflect the beam, and thus the coupling efficiencies into the two fibres leading to the detectors might change between measurement stage a and b. I found that when coupling into multimode (instead of single mode) fibre in the output ports of this beamsplitter, there was no measurable change in coupling efficiency.

Finally, measurement stages a and b should be performed back-to-back to minimize the chance of the source power drifting significantly over the time of the measurement.

C.3 Double pairs

SPDC sources sometimes produce two photon pairs, instead of one. If the ratio of double pairs to single pairs is too high this will add additional degrees of freedom to the data. The

cause of this can be explained by a bucket detector's nonlinear response to the number of incident photons [16].

A detector with efficiency η will click with probability $p(\text{click}|\bullet) = \eta$ when a single photon is incident on it. If two photons ($\bullet\bullet$) are incident on the bucket detector, the leading photon will make the detector click with probability η, and, in the event that the first photon *doesn't* make the detector click (which happens with probability $1-\eta$), the second photon has a chance η to make the detector click. This leads to a total clicking probability of [16] $p(\text{click}|\bullet\bullet) = \eta + (1-\eta)\eta = 2\eta - \eta^2$. The expected number of counts recorded by the detector, n, is a function of these probabilities:

$$n_{\text{clicks}} \propto p(\text{click}|\bullet)p(\bullet) + p(\text{click}|\bullet\bullet)p(\bullet\bullet). \tag{C.2}$$

For a heralded SPDC source, $p(\bullet\bullet) = |\gamma|^2 p(\bullet)$, and the expected number of counts reduces to:

$$n \propto \eta + |\gamma|^2(2\eta - \eta^2). \tag{C.3}$$

For light that travels through a polarizing beamsplitter, the number of counts recorded in each output arm of the beamsplitter are:

$$n_t \propto \eta_t + |\gamma|^2(2\eta_t - \eta_t^2), \tag{C.4}$$
$$n_r \propto \eta_r + |\gamma|^2(2\eta_r - \eta_r^2), \tag{C.5}$$
$$\tag{C.6}$$

where I defined $\eta_t = T\eta$ as the probability, T, that a photon is transmitted at the beamsplitter multiplied by the the detector efficiency η, and I defined $\eta_r = (1-T)\eta$ in a similar way. (I'll assume for now that the both detectors have the same efficiency η and that there is no need to perform the measurement in two stages). Thus the probability of obtaining outcome '0' is:

$$p(0) = \frac{n_t}{n_t + n_r} = \frac{T\eta + |\gamma|^2(2T\eta - T^2\eta^2)}{T\eta + |\gamma|^2(2T\eta - T^2\eta^2) + (1-T)\eta + |\gamma|^2(2(1-T)\eta - (1-T)^2\eta^2)}, \tag{C.7}$$

which is a function of T, η, and $|\gamma|^2$. Note that if the detectors responded *linearly* to photon number (i.e. if $p(\text{click}|\bullet\bullet) = 2\eta$), then Eq. (C.7) would reduce to $p(0) = T$, which is what we desire. However, because of the detector nonlinearity, double pairs of photons produce different statistics than single pairs.

I ran my experiments at low power, with a measured heralded $g_h^{(2)}(0) \lesssim 0.002$, which implies a double-to-single pair ratio of $\approx 1:2000$.

When setting up the experiment in Chapter 5, I noticed that timing settings on the coincidence electronics was important. For example, increasing the coincidence window from 2 ns to 3 ns increased the apparent number of degrees of freedom in my data. The data was also sensitive to the timing delay between the herald detector and the two other detectors. For example, when using a coincidence window of 1 ns, I was able to find a rank-4 GPT that fit the data "well", for one timing delay setting. I then changed the delay of the herald detector by ≈ 250 ps (while changing nothing else about the experiment), and the rank-4 GPT fit to the resulting data was very poor.

I am not completely sure why the dimension of the data was so sensitive to the settings of the coincidence electronics, but my hypothesis is that both increasing the coincidence window and using an imperfect timing delay increases the number of double pairs detected relative to the number of single pairs.

C.4 XPS-controlled rotation mounts

The XPS-controlled waveplate mounts from Newport have a shiny side and a dull side. The shiny side has four *really* shiny screws in the rotating part of the mount, the heads of which reflect background light into your detectors. These should be covered up with black tape.