Experimentally bounding deviations from quantum theory for a photonic three-level system using theory-agnostic tomography

Michael J. Grabowecky [1,*] Christopher A. J. Pollack,[1,2] Andrew R. Cameron,[1] Robert W. Spekkens [3] and Kevin J. Resch[1]

[1]*Institute for Quantum Computing and Department of Physics & Astronomy, University of Waterloo, Waterloo, Ontario, N2L 3G1, Canada*
[2]*Department of Mathematics, University of Toronto, Toronto, Ontario, M5S 1A1, Canada*
[3]*Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, N2L 2Y5*

If one seeks to test quantum theory against many alternatives in a landscape of possible physical theories, then it is crucial to be able to analyze experimental data in a theory-agnostic way. This can be achieved using the framework of generalized probabilistic theories (GPTs). Here we implement GPT tomography on a three-level system corresponding to a single photon shared among three modes. This scheme achieves a GPT characterization of each of the preparations and measurements implemented in the experiment without requiring any prior characterization of either. Assuming that the sets of realized preparations and measurements are tomographically complete, our analysis identifies the most likely dimension of the GPT vector space describing the three-level system to be nine, in agreement with the value predicted by quantum theory. Relative to this dimension, we infer the scope of GPTs that are consistent with our experimental data by identifying polytopes that provide inner and outer bounds for the state and effect spaces of the true GPT. From these, we are able to determine quantitative bounds on possible deviations from quantum theory. In particular, we bound the degree to which the no-restriction hypothesis might be violated for our three-level system.

I. INTRODUCTION

Despite the fact that quantum theory has thus far been very successful in describing nature, it may one day be superseded by a novel postquantum theory. In recent years, there have been many theoretical milestones in identifying plausible candidates for such a theory [1–3]. Some of these theories can even be explained within the quantum formalism itself, such as models that exhibit intrinsic decoherence [4–7]. However, most theories require a rejection of some or all of the quantum framework. Some important examples include Almost Quantum Theory [2,8], theories involving higher order interference [9–12], and theories involving quaternions [13,14].

In contrast to the abundance of theoretical proposals, there have been few experiments that aim to test quantum theory simultaneously against many alternative physical theories. Doing so is difficult as one cannot *a priori* assume the validity of any one particular theory. For this reason, a theory-neutral framework must be adopted to analyze the experimental data. Such a framework is provided by the formalism of *generalized probabilistic theories* (GPTs).

The GPT formalism provides a description of a physical system from an operational approach [1,15–23] and is used frequently in the field of quantum foundations. It is operational because it describes the theory based solely on what it predicts for the probabilities for each outcome of a measurement in an experiment. This framework requires the use of two weak assumptions, both of which are adopted in standard

quantum theory. The first is that each choice of preparation is made independently from each choice of measurement. The other assumption is that repeated runs of the same experiment yield an independent and identically distributed (i.i.d.) source of data. Under the validity of these two assumptions, the GPT framework is completely general for modeling experimental data, allowing one to avoid any implicit bias towards quantum theory.

Recent work placed bounds on possible deviations from quantum theory within a landscape of alternative theories for a two-level system using the GPT framework [24]. In that work, the two-level system was the polarization degree of freedom of a single photon. A large number of repetitions of an experiment on photon polarization were conducted, across which the preparation procedure and the measurement procedure was varied (from among a large set of possibilities for each). The authors developed a scheme for achieving a GPT characterization of these preparations and measurements, that is, a GPT tomography scheme. This scheme requires no prior characterization of the preparations or measurements. Rather, these are simultaneously characterized based on their interplay. We will refer to this scheme as *bootstrap GPT tomography*. Bootstrap GPT tomography involves *two* key innovations relative to standard quantum tomography [25,26]. First, standard quantum tomography schemes are valid only under the assumption of the correctness of quantum theory. They do not return *GPT* characterizations of the states and measurements, but only *quantum* characterizations thereof. In this sense, they fail to be *theory-agnostic*. Second, standard schemes achieve a characterization of the states under the assumption of a prior characterization of the measurements [25], or they achieve a

*mgrabowecky@uwaterloo.ca

characterization of the measurements under the assumption of a prior characterization of the states [26]. They do not achieve a *bootstrap* characterization of these, that is, a simultaneous characterization of both states and measurements based on their interplay.

In the present work, we apply bootstrap GPT tomography to a photonic three-level system. The constraints describing the geometry of the quantum state and effect spaces of three-level systems are more complex when compared with their two-level counterparts [27]. Due to this complexity, there is a greater opportunity for an experiment to reveal some surprising discrepancies between nature and the predictions of quantum theory [28].

Our experiment is conducted using heralded single photons prepared using spontaneous parametric down conversion (SPDC). Three-level states are encoded in three modes of single photons where the modes are distinguished by a combination of their polarization and spatial degrees of freedom. We perform bootstrap tomography on a large set of preparations and a large set of measurements to obtain a GPT characterization of both. This bootstrap tomography approach differs from other experimental schemes that aim to characterize the geometry of three-level photonic systems, for example the one outlined in Ref. [29], in that the correctness of the quantum formalism is not presumed in advance.

Our analysis uses a model selection technique in which data is collected for two independent runs of the experiment. The first data set is called the *training set,* and the second data set is called the *test set*. For each of a set of candidate dimensions, we find the GPT characterization that best fits the training set. We then score these characterizations based on how well they predict the test set. This model selection technique was previously applied to the analysis of a quantum experiment in Daley *et al.* [30], where the goal was to adjudicate between different causal accounts of a Bell experiment. Here it is applied to the problem of achieving bootstrap GPT tomography.

In the original version of the bootstrap GPT tomography scheme, described in Ref. [24], it is the Akaïke information criterion rather than a train-and-test technique that is used to adjudicate between models of different dimensions. Although the Akaïke information criterion has the advantage of yielding an estimate of the probability that a given model in the slate of candidates minimizes the estimated information loss relative to the true model, it does so only under strong assumptions about the error model underlying the experimental data, and these assumptions do not, strictly speaking, hold in our experiment. Similar considerations hold for some of the proposed alternatives to the Akaïke information criterion (see, e.g., [31,32]). An additional drawback of model selection techniques based on the Akaïke information criterion and other criteria like it is that they are typically reliable only in situations where the number of data points is large compared to the number of parameters. As this is not the situation in our experiment, we could not be confident of the verdicts delivered by such criteria. The train-and-test technique, although more heuristic than those based on information criteria, can be applied regardless of the error model and the relative number of data points and parameters. It is for these reasons that we have opted to use it here.

We now summarize the results obtained from our experimental data analysis and what can be inferred regarding possible deviations from quantum theory. (For a more in-depth discussion of what one can infer about possible discrepancies in the dimension and shapes of the GPT state and effect spaces from experiments such as the one described here, we refer the reader to the introduction of Ref. [24].)

There are two ways in which the true GPT describing the three-level system may differ from quantum-mechanical predictions. First, the experimental data might imply a deviation in the *dimension* of the GPT vector space (the number of parameters needed to specify a state or effect in the theory) relative to quantum theory. Second, the experimental data might imply no dimensional deviation, but a deviation in the *shape* of the state and effect spaces relative to quantum theory.

Because we are not assuming the correctness of quantum theory, the set of preparations and measurements that we implement might fail to be tomographically complete for the three-level system of interest. Furthermore, the additional preparations and measurements required to achieve tomographic completeness might be implementable only in a distinct experimental setup from the one we consider, in particular, one involving exotic physics. Because of this, a dimensional discrepancy between the true GPT and quantum theory might exist but be missed by our experiment. Nonetheless, if a dimensional discrepancy exists, it is possible that it could be detected in a nonexotic experimental setup (such as ours) because evidence for exotic physics can arise in conventional setups when these are probed at the precision frontier. In this sense, our experiment provides an opportunity for uncovering such a discrepancy.

Applying our train-and-test model selection technique to the experimental data, we determine that the most likely dimension of the GPT vector space governing our three-level system is nine, in agreement with the dimension predicted by qutrit quantum theory. The fact that we find no evidence in favor of the hypothesis of a dimensional discrepancy implies that one of the following possibilities must hold: (1) it does not exist; (2) it exists, but can be detected only in an experimental setup involving exotic physics; or (3) it exists and can be detected in the sort of experiment we have implemented but only at higher precision than we achieved here.

Under the assumption that there is no dimensional discrepancy, we deduce from our experimental data the geometry of the spaces of experimentally realized states and effects, and from these we deduce the geometry of the spaces of logically consistent states and effects. Together, these yield inner and outer bounding polytopes for the state and effect spaces of the true GPT governing the three-level system. The bounds we obtain do not exclude the state and effect spaces of quantum theory, and therefore our results are consistent with the hypothesis that quantum theory is the true GPT underlying our experiment. Furthermore, using these bounding polytopes on the state and effect spaces, we are able to derive a quantitative bound on the extent to which the true GPT might violate the no-restriction hypothesis [16]. In Ref. [24] this was done by estimating the ratio of the volumes of the inner and outer bounding polytopes for the state space. In the higher-dimensional context considered here, however, computing these volumes is infeasible, and so we instead estimate

the ratio of the distance from the origin to the inner and outer bounding polytopes for the state space, averaged over a large number of directions. This linear dimension ratio was found to be 0.80 ± 0.04.

In summary, our analysis is consistent with quantum theory being correct. If the true GPT *does* involve a dimensional deviation from quantum theory, then detecting it will require either achieving higher precision in a conventional setup of the type we implemented, or an experimental setup involving exotic physics. Similarly, if the true GPT involves no dimensional deviation, but a deviation in the shapes of the state and effect spaces, then any future experiment establishing this will find such deviations to be limited in scope by the quantitative extent established here. Our experiment and others of its kind thereby provide an empirical constraint—a kind of stress test—that can be applied to any novel theory that is proposed as a competitor to quantum theory.

II. BASICS OF THE GPT FRAMEWORK

A GPT describes an experiment operationally by predicting the probability distribution over outcomes for every measurement given every preparation. To characterize the set of effects appearing in a GPT, it is sufficient to consider binary-outcome measurements, since an effect that appears in an N-outcome measurement also appears in the binary-outcome measurement that coarse grains all the outcomes that are not associated with that effect. Furthermore, for any binary-outcome measurement, it is sufficient to specify the probability of only one of its outcomes, say, the 0 outcome, since by normalization the probability of the other is then fixed. Thus, for each of a set of m preparations $\{P_i\}_{i=1}^m$ and a set of n binary-outcome measurements $\{M_j\}_{j=1}^n$, the GPT need specify only the probability of obtaining the 0 outcome, denoted $p(0|P_i, M_j)$. The full set of probabilities obtained from each of the preparation-measurement pairs in an experiment can be organized into an $m \times n$ matrix D, where each row represents a preparation, and each column represents a measurement:

$$D = \begin{pmatrix} p(0|P_1, M_1) & p(0|P_1, M_2) & \cdots & p(0|P_1, M_n) \\ p(0|P_2, M_1) & p(0|P_2, M_2) & \cdots & p(0|P_2, M_n) \\ \vdots & \vdots & \ddots & \vdots \\ p(0|P_m, M_1) & p(0|P_m, M_2) & \cdots & p(0|P_m, M_n) \end{pmatrix}.$$

The matrix D encapsulates all of the GPT's predictions about the preparations and measurements in the experiment. If D is rank k, then one can perform the decomposition $D = SE$, where S is an $m \times k$ matrix and E is a $k \times n$ matrix. The rows of S are called *GPT state vectors* and are denoted s_i, $i \in \{1, 2, \ldots, m\}$. The columns of E are called *GPT effect vectors* and are denoted e_j, $j \in \{1, 2, \ldots, n\}$. Therefore, each element of D can be written as

$$D_{i,j} = p(0|P_i, M_j) = \boldsymbol{s}_i \cdot \boldsymbol{e}_j. \tag{1}$$

A GPT is fully specified by the set of all physically realizable state vectors, termed the *GPT state space* and denoted $\boldsymbol{\mathcal{S}}$, as well as the set of all physically realizable effect vectors, termed the *GPT effect space* and denoted $\boldsymbol{\mathcal{E}}$. The decomposition of D is not unique. Specifically, for any invertible $k \times k$ matrix Λ, one can write $D = SE = S\Lambda\Lambda^{-1}E$. For each

choice of Λ, the rows of the matrix $S\Lambda$ consist of a set of valid GPT state vectors and the columns of the matrix $\Lambda^{-1}E$ consist of a set of valid GPT effect vectors. For any two such decompositions, therefore, the GPT state and effect spaces defined by one are linearly related to those defined by the other. There is consequently a linear freedom in the GPT state and effect spaces, and stipulating a particular choice of linear transformation is therefore just a convention.

GPTs are typically formulated such that any convex combination of preparation (measurement) procedures is also a valid preparation (measurement) procedure. This is justified by the requirement of including classical probability theory as a subtheory of the GPT, which allows for arbitrary mixing and postprocessing. Consequently, the GPT state and effect spaces are convex. In addition, it is assumed that for any valid GPT effect vector $\boldsymbol{e}_j$ in the set $\boldsymbol{\mathcal{E}}$, the vector $\bar{\boldsymbol{e}}_j = \boldsymbol{u} - \boldsymbol{e}_j$ is also a valid effect vector, where the vector $\boldsymbol{u}$ is called the unit effect and assigns probability 1 to all states. We adopt the convention that $\boldsymbol{u} = (1, 0, \ldots, 0)^T$, which implies that the first component of every normalized state must be 1, or equivalently, that the first column of the matrix S contains only ones. We also adopt the convention that the effect with index $j = 1$ is the unit effect, so that the first column of the matrix D also contains only ones. Thus, each k-dimensional state vector is fully described by a set of $k - 1$ parameters.

We call a vector $\boldsymbol{s}$ a *logically possible state* if it assigns a valid probability to every measurement effect allowed by the GPT. More formally, we define the space of logically possible states, denoted $\boldsymbol{\mathcal{S}}^{\text{logical}}$, as the set of states $\boldsymbol{s}$ such that $\forall \boldsymbol{e} \in \boldsymbol{\mathcal{E}}: 0 \leqslant \boldsymbol{s} \cdot \boldsymbol{e} \leqslant 1$ and $\boldsymbol{s} \cdot \boldsymbol{u} = 1$. From its definition, it can be shown that $\boldsymbol{\mathcal{S}}^{\text{logical}}$ is the intersection of the geometric dual of $\boldsymbol{\mathcal{E}}$ and the hyperplane $\boldsymbol{s} \cdot \boldsymbol{u} = 1$. For simplicity, we write $\boldsymbol{\mathcal{S}}^{\text{logical}} = \text{dual}(\boldsymbol{\mathcal{E}})$. In an analogous manner, we define the set of *logically possible effects*, denoted $\boldsymbol{\mathcal{E}}^{\text{logical}}$, as the set of all vectors $\boldsymbol{e}$ such that $\forall \boldsymbol{s} \in \boldsymbol{\mathcal{S}}: 0 \leqslant \boldsymbol{s} \cdot \boldsymbol{e} \leqslant 1$. Again, for simplicity, we write $\boldsymbol{\mathcal{E}}^{\text{logical}} = \text{dual}(\boldsymbol{\mathcal{S}})$.

GPTs in which $\boldsymbol{\mathcal{S}}^{\text{logical}} = \boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{E}}^{\text{logical}} = \boldsymbol{\mathcal{E}}$ are said to satisfy the *no-restriction hypothesis* [16]. For GPTs which satisfy the no-restriction hypothesis, all logically allowed GPT states are physically allowed and similarly for the effects. It is not possible for an experiment to prove that a particular GPT satisfies the no-restriction hypothesis exactly. This is because the number of states and effects characterized in an experiment will always be finite, and noise in their implementation will bound the realized GPT spaces away from the theoretical ideal. Therefore, the best one can do is constrain the extent to which any GPT that is logically consistent with the experimental data can violate the no-restriction hypothesis.

III. THE GPT STATE AND EFFECT SPACES DESCRIBING A QUTRIT

One of the most familiar examples of a GPT is quantum mechanics itself. Here we will consider qutrit quantum mechanics. In this case, the GPT state space $\boldsymbol{\mathcal{S}}^{\text{qutrit}}$ is isomorphic to the set of all positive semidefinite trace-one linear operators acting on a three-dimensional complex vector space, i.e., the set $\text{QStates}(\mathbb{C}^3) = \{\rho \in \mathcal{L}(\mathbb{C}^3): \rho \geqslant 0, \text{Tr}(\rho) = 1\}$. The GPT effect space $\boldsymbol{\mathcal{E}}^{\text{qutrit}}$ is isomorphic to the set of all positive semidefinite linear operators with eigenvalues less than

or equal to one, i.e., the set $\mathrm{QEffs}(\mathbb{C}^3) = \{Q \in \mathcal{L}(\mathbb{C}^3) : Q \geqslant 0, Q \leqslant I\}$.

A particularly useful convention to adopt for the GPT representation of the spaces of qutrit quantum states and quantum effects is to use the generalized Bloch representation. This representation is constructed from the identity operator together with a particular basis of the eight-dimensional vector space of traceless Hermitian operators (which is orthonormal relative to the Hilbert-Schmidt inner product), namely, the basis consisting of the eight Gell-Mann matrices [33]. These matrices generate the Lie algebra of the group SU(3), in analogy with how the Pauli matrices generate the Lie algebra of the group SU(2). Denoting the Gell-Mann matrices by $\{\lambda_\alpha\}_{\alpha=1}^8$, the nine operators used in the Bloch representation are

$$
\lambda_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \lambda_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},
$$

$$
\lambda_2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \lambda_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},
$$

$$
\lambda_4 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \qquad \lambda_5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix}, \qquad (2)
$$

$$
\lambda_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad \lambda_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix},
$$

$$
\lambda_8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix}.
$$

One easily verifies that the Gell-Mann matrices are Hermitian, traceless ($\mathrm{Tr}[\lambda_\alpha] = 0 \ \forall \alpha \neq 0$), and orthonormal relative to the Hilbert-Schmidt inner product on the operator space ($\mathrm{Tr}[\lambda_\alpha \lambda_\beta] = 2\delta_{\alpha\beta} \ \forall \alpha, \beta \neq 0$, where the normalization is taken to be 2). With this notation, the Bloch representation of any state $\rho \in \mathrm{QStates}(\mathbb{C}^3)$ and any effect $Q \in \mathrm{QEffs}(\mathbb{C}^3)$ is given by

$$
\rho = \frac{1}{3}I + \frac{1}{2}\sum_{\alpha=1}^8 s_\alpha \lambda_\alpha
$$

and

$$
Q = e_0 I + \sum_{\alpha=1}^8 e_\alpha \lambda_\alpha,
$$

respectively. The GPT state vectors are given by the Bloch coefficients of ρ: $\boldsymbol{s} = (1, s_1, \ldots, s_8)$, where $s_\alpha = \mathrm{Tr}[\rho \lambda_\alpha] \ \forall \alpha \in \{0, \ldots, 8\}$. Furthermore, the GPT effect vectors are given by the Bloch coefficients of Q: $\boldsymbol{e} = (e_0, e_1, \ldots, e_8)$ where $e_\alpha = \frac{1}{2}\mathrm{Tr}[Q\lambda_\alpha] \ \forall \alpha \in \{0, \ldots, 8\}$. It follows that $p(0|P_i, M_j) = \mathrm{Tr}[\rho_i Q_j] = \boldsymbol{s}_i \cdot \boldsymbol{e}_j$, thereby matching Eq. (1). Note that in this generalized Bloch representation, the quantum unit effect, I, is represented by the GPT effect vector having $e_0 = 1$ and $e_\alpha = 0, \forall \alpha \neq 0$, so it abides by the convention that the unit effect vector is taken to be $\boldsymbol{u} = (1, 0, \ldots, 0)$.

Let $\tilde{\boldsymbol{s}} = (s_1, \ldots, s_8)$ be the eight-dimensional vector obtained from $\boldsymbol{s}$ by eliminating its first component, and let $||\tilde{\boldsymbol{s}}|| = \sqrt{s_1^2 + s_2^2 + \cdots + s_8^2}$. It has been shown that $\boldsymbol{s}$ is a GPT state vector for a valid qutrit state in the generalized Bloch representation ($\boldsymbol{s} \in \mathcal{S}^{\mathrm{qutrit}}$) if and only if it satisfies the following constraints [34]:

$$
\begin{aligned}
&||\tilde{\boldsymbol{s}}|| \leqslant \frac{2}{\sqrt{3}}, \\
&\frac{2}{9} - \frac{1}{2}||\tilde{\boldsymbol{s}}||^2 + \frac{1}{2}\sum_{\alpha,\beta,\gamma} g_{\alpha\beta\gamma} s_\alpha s_\beta s_\gamma \geqslant 0,
\end{aligned} \qquad (3)
$$

where $g_{\alpha\beta\gamma}$ are the elements of the completely symmetric structure tensor of SU(3). The elements of g are defined implicitly via the anticommutation relations of the Gell-Mann matrices [34,35]:

$$
\{\lambda_\alpha, \lambda_\beta\} = \frac{4}{3}\delta_{\alpha\beta}I + 2g_{\alpha\beta\gamma}\lambda_\gamma, \quad \forall \alpha, \beta, \gamma \neq 0. \qquad (4)
$$

The first constraint in Eq. (3) corresponds to inclusion in an eight-dimensional hypersphere, which is the qutrit analog of the well-known Bloch sphere representation of qubits. The second constraint in Eq. (3) has no qubit analog and makes the qutrit state space geometry much more complex.

Defining $\tilde{\boldsymbol{e}} = (e_1, \ldots, e_8)$ and $||\tilde{\boldsymbol{e}}|| = \sqrt{e_1^2 + \cdots + e_8^2}$, we have that $\boldsymbol{e}$ is the GPT effect vector for a valid qutrit effect in the generalized Bloch representation ($\boldsymbol{e} \in \mathcal{E}^{\mathrm{qutrit}}$) if and only if it satisfies the following constraints:

$$
\begin{aligned}
&0 \leqslant e_0 \leqslant 1, \\
&||\tilde{\boldsymbol{e}}|| \leqslant \sqrt{3}\mathrm{Min}(e_0, \ 1 - e_0), \\
&e_0^3 - e_0||\tilde{\boldsymbol{e}}||^2 + \frac{2}{3}\sum_{\alpha,\beta,\gamma} g_{\alpha\beta\gamma} e_\alpha e_\beta e_\gamma \geqslant 0, \\
&(1 - e_0)^3 - (1 - e_0)||\tilde{\boldsymbol{e}}||^2 - \frac{2}{3}\sum_{\alpha,\beta,\gamma} g_{\alpha\beta\gamma} e_\alpha e_\beta e_\gamma \geqslant 0.
\end{aligned} \qquad (5)
$$

The constraints in Eq. (5) are derived following the same steps as are followed to derive Eq. (3) [34], with the added condition that $Q \leqslant I$. Note that the third and fourth conditions follow from $Q \geqslant 0$ and $I - Q \geqslant 0$, respectively.

To visualize the geometry of the nine-dimensional qutrit state and effect spaces, it is useful to consider the geometry of various three-dimensional projections of these. We follow the approach taken in Ref. [36] to the study of the geometry of the qutrit state space (see also Refs. [37–39]) and adapt this approach to the case of the effect space.

The *joint numerical range* of a set of k operators $\{A_\alpha\}_{\alpha=0}^{k-1}$ is defined as [36]

$$
\begin{aligned}
W(A_0, \ldots, A_{k-1}) = \{\boldsymbol{w} \in \mathbb{R}^k : w_\alpha = \langle\psi|A_\alpha|\psi\rangle, \\
\forall|\psi\rangle \in \mathbb{C}^d\}, \qquad (6)
\end{aligned}
$$

where $\mathbb{C}^d$ is the complex vector space of dimension d. Consider the nine operators defined by

$$
A_0 = I, \ A_\alpha = \lambda_\alpha, \ \forall \alpha \in \{1, \ldots, 8\},
$$

i.e., the identity matrix and the eight Gell-Mann matrices, defined in Eq. (2). The joint numerical range $W(A_0, \ldots, A_8)$ is equal to the set of nine-dimensional generalized Bloch vectors associated with pure qutrit states (i.e., the Bloch vectors

including the s_0 component),

$$\mathcal{S}_{\text{pure}}^{\text{qutrit}} = W(A_0, \ldots, A_8). \tag{7}$$

It follows that the *convex hull* of $W(A_0, \ldots, A_8)$ is equal to $\mathcal{S}^{\text{qutrit}}$, which corresponds to the set of nine-dimensional generalized Bloch vectors associated with *all* qutrit states, characterized in Eq. (3),

$$\mathcal{S}^{\text{qutrit}} = \{s \in \mathbb{R}^9 : s_\alpha = \text{Tr}[\rho A_\alpha] \, \forall \alpha \in \{0, \ldots, 8\},$$
$$\forall \rho \in \text{QStates}(\mathbb{C}_3)\},$$
$$= \text{cvxhull}[W(A_0, \ldots, A_8)]. \tag{8}$$

For some arbitrary triple of these nine operators, $\{A_\mu, A_\nu, A_\omega\}$, the convex hull of their joint numerical range describes the *projection* of $\mathcal{S}^{\text{qutrit}}$ into the three-dimensional operator subspace spanned by $\{A_\mu, A_\nu, A_\omega\}$,

$$\mathcal{P}_{\mu\nu\omega}(\mathcal{S}^{\text{qutrit}}) = \{s \in \mathbb{R}^3 : s_\alpha = \text{Tr}[\rho A_\alpha], \forall \alpha \in \{\mu, \nu, \omega\},$$
$$\forall \rho \in \text{QStates}(\mathbb{C}^3)\}$$
$$= \text{cvxhull}[W(A_\mu, A_\nu, A_\omega)]. \tag{9}$$

Similar considerations hold for the effects. Taking

$$B_0 = \tfrac{1}{3}I, \quad B_\alpha = \tfrac{1}{2}\lambda_\alpha, \; \forall \alpha \in \{1, \ldots, 8\},$$

the joint numerical range $W(B_0, \ldots, B_8)$ is equal to the set of generalized Bloch vectors associated with *atomic* effects on qutrits. An atomic quantum effect is one described by a rank-1 projector, so that the set of atomic quantum effects for a qutrit is simply $\text{AtmQEffs}(\mathbb{C}^3) = \{|\psi\rangle\langle\psi| : |\psi\rangle \in \mathbb{C}^3\}$. Denoting the set of GPT effect vectors that are associated with atomic qutrit effects by $\mathcal{E}_{\text{atomic}}^{\text{qutrit}}$, we have

$$\mathcal{E}_{\text{atomic}}^{\text{qutrit}} = W(B_0, \ldots, B_8). \tag{10}$$

For all such GPT effect vectors, $e_0 = \text{Tr}(\tfrac{1}{3}|\psi\rangle\langle\psi|) = \tfrac{1}{3}$.

To transition from the set of atomic quantum effects to the set of *all* quantum effects, the following procedure suffices. For each atomic quantum effect $Q \in \text{AtmQEffs}(\mathbb{C}^3)$, add to the set its complement, $I - Q$. In addition, add to the set the unit effect, I, and the zero effect, 0, which assigns probability 0 when paired with all states. Finally, close under convex combinations. In summary, if one denotes the set of complements of atomic quantum effects by $\overline{\text{AtmQEff}}(\mathbb{C}^3)$, then the full set of quantum effects satisfies

$$\text{QEff}(\mathbb{C}^3) = \text{cvxhull}\big[\{0\} \cup \{I\} \cup \text{AtmQEff}(\mathbb{C}^3)$$
$$\cup \overline{\text{AtmQEff}}(\mathbb{C}^3)\big]. \tag{11}$$

The same procedure allows one to go from the set of GPT effect vectors describing atomic qutrit effects in the generalized Bloch representation to the set of GPT effect vectors describing generic qutrit effects. Recall that $\mathbf{u} = (1, 0, \ldots, 0)$ denotes the GPT effect vector associated with the unit quantum effect. Letting $\mathbf{0} = (0, \ldots, 0)$ denote the GPT effect vector associated with the zero quantum effect and letting $\bar{\mathcal{E}}_{\text{atomic}}^{\text{qutrit}}$ denote the complement of the set of atomic effects,

$\bar{\mathcal{E}}_{\text{atomic}}^{\text{qutrit}} = \{\mathbf{u} - \mathbf{e} : \mathbf{e} \in \mathcal{E}_{\text{atomic}}^{\text{qutrit}}\}$, we have

$$\mathcal{E}^{\text{qutrit}} = \{e \in \mathbb{R}^9 : e_\alpha = \text{Tr}[QB_\alpha] \, \forall \alpha \in \{0, \ldots, 8\},$$
$$\forall Q \in \text{QEff}(\mathbb{C}^3)\}$$
$$= \text{cvxhull}\big[\{\mathbf{0}\} \cup \{\mathbf{u}\} \cup \mathcal{E}_{\text{atomic}}^{\text{qutrit}} \cup \bar{\mathcal{E}}_{\text{atomic}}^{\text{qutrit}}\big]. \tag{12}$$

As with the states, we wish to consider the *projection* of $\mathcal{E}^{\text{qutrit}}$ into the three-dimensional operator subspace spanned by $\{B_\mu, B_\nu, B_\omega\}$ for various triples of operators. Some preliminary definitions are useful for describing this projection. Let $\mathbf{0}^{(3)} \equiv (0, 0, 0)$; this is the three-dimensional projection of $\mathbf{0}$ for any choice of μ, ν, ω. Let $\mathbf{u}^{(\mu\nu\omega)}$ denote the projection of $\mathbf{u}$ into the appropriate three-dimensional subspace. Note that if $0 \in \{\mu, \nu, \omega\}$, then the corresponding component of $\mathbf{u}^{(\mu\nu\omega)}$ is 1, with the other components being 0, while if $0 \notin \{\mu, \nu, \omega\}$, then $\mathbf{u}^{(\mu\nu\omega)} = \mathbf{0}^{(3)}$. Finally, let $\overline{W}(B_\mu, B_\nu, B_\omega)$ denote the complement of $W(B_\mu, B_\nu, B_\omega)$ relative to $\mathbf{u}_{\mu\nu\omega}$, i.e., $\overline{W}(B_\mu, B_\nu, B_\omega) = \{\mathbf{u}^{(\mu\nu\omega)} - \mathbf{e} : \mathbf{e} \in W(B_\mu, B_\nu, B_\omega)\}$. For those projections where $\mathbf{u}^{(\mu\nu\omega)} = \mathbf{0}^{(3)}$, $\overline{W}(B_\mu, B_\nu, B_\omega)$ is simply the inversion about the origin of $W(B_\mu, B_\nu, B_\omega)$.

Relative to these definitions, the three-dimensional projections of $\mathcal{E}^{\text{qutrit}}$ are given by

$$\mathcal{P}_{\mu\nu\omega}(\mathcal{E}^{\text{qutrit}}) = \{e \in \mathbb{R}^3 : e_\alpha = \text{Tr}[QB_\alpha], \forall Q \in \text{QEff}(\mathbb{C}^3)\}$$
$$= \text{cvxhull}[\{\mathbf{0}^{(3)}\} \cup \{\mathbf{u}^{(\mu\nu\omega)}\}$$
$$\cup W(B_\mu, B_\nu, B_\omega) \cup \overline{W}(B_\mu, B_\nu, B_\omega)]. \tag{13}$$

Figure 1 shows plots of a number of three-dimensional projections of the qutrit state space (yellow shapes) and the qutrit effect space (cyan shapes). Before describing some of their features, we briefly explain how these plots are generated.

To plot a given three-dimensional projection of $\mathcal{S}^{\text{qutrit}}$, we begin by generating a large quantity of random GPT state vectors that are on the boundary of the set defined by Eq. (3) [40]. For every GPT state vector in our random set, we project it down into the three-dimensional space of interest. We then plot the convex hull of these three-dimensional vectors. Similarly, to generate a plot of any given three-dimensional projection of $\mathcal{E}^{\text{qutrit}}$, we generate a large quantity of random GPT effect vectors that are on the boundary of the set defined by Eq. (5) and with $e_0 = \tfrac{1}{3}$, which is to say, vectors associated with atomic qutrit effects. For every GPT effect vector obtained in this way, we compute its complement effect. We then project both the atomic effects and their complements down into the three-dimensional space of interest. Finally, we plot the convex hull of the set of projected GPT effect vectors, their projected complements, and the projected zero and unit effect vectors. As the number of vectors defining these convex shapes is always finite, the shapes depicted in the plots are polytopes. Indeed, a close examination of Fig. 1 reveals that they are not smooth, in contrast to the true three-dimensional projections. Nonetheless, because the sampling is quite dense, the plots provide a good visual approximation to the true shapes.

Consider first the three-dimensional projections of the nine-dimensional qutrit state space onto a triple of Gell-Mann matrices, that is, the projections where $0 \notin \{\mu, \nu, \omega\}$. There are $\binom{8}{3} = 56$ of these. In Ref. [38,39] the connection to joint numerical ranges was leveraged to show that all
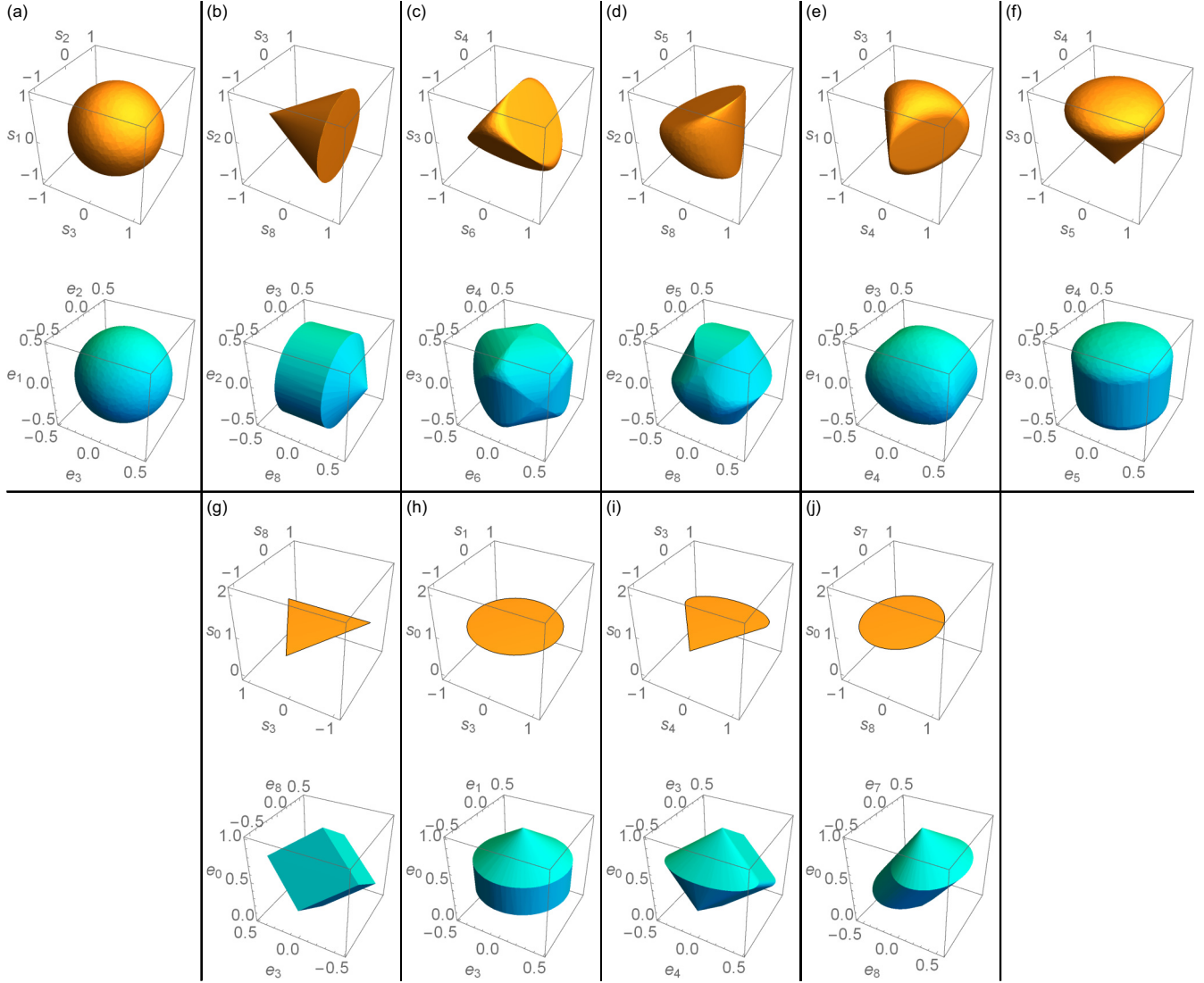
FIG. 1. A sampling of three-dimensional projections of the nine-dimensional qutrit state space (top shape of each subfigure, displayed in yellow) and of the nine-dimensional qutrit effect space (bottom shape of each subfigure, displayed in cyan). For each subfigure (a)–(j), we specify the triple of indices (μ, ν, ω) describing which three-dimensional projection is plotted, as well as the type of the state space projection. In the cases (a)–(f), the type is a pair of numbers (l, f), counting the flat segments and flat faces in the two-dimensional boundary, while in the cases (g)–(f), it is a single number l counting the flat segments in the one-dimensional boundary. (a) $(\mu, \nu, \omega) = (1, 2, 3)$, $(l, f) = (0, 0)$; (b) $(\mu, \nu, \omega) = (2, 3, 8)$, $(l, f) = (\infty, 1)$; (c) $(\mu, \nu, \omega) = (3, 4, 6)$, $(l, f) = (0, 4)$; (d) $(\mu, \nu, \omega) = (1, 3, 4)$, $(l, f) = (0, 2)$; (e) $(\mu, \nu, \omega) = (2, 5, 8)$, $(l, f) = (1, 2)$; (f) $(\mu, \nu, \omega) = (3, 4, 5)$, $(l, f) = (\infty, 0)$; (g) $(\mu, \nu, \omega) = (0, 3, 8)$, $l = 3$; (h) $(\mu, \nu, \omega) = (0, 1, 3)$, $l = 0$; (i) $(\mu, \nu, \omega) = (0, 3, 4)$, $l = 2$; (j) $(\mu, \nu, \omega) = (0, 7, 8)$, $l = 0$.

three-dimensional projections of the qutrit state space can be classified into different types according to the number of *flat segments* in their two-dimensional boundary and the number of *flat faces* in their two-dimensional boundary. A given type of three-dimensional projection, therefore, can be specified by a pair of natural numbers, denoted (l, f), where l is the number of flat segments and f is the number of flat faces.

Figures 1(a)–1(f) depicts a sampling of three-dimensional projections of the qutrit state space, corresponding to six distinct types. The projection depicted in Fig. 1(a), for example, is of type $(l, f) = (0, 0)$ because it is a ball and therefore has no flat segments or flat faces, while for Fig. 1(e), the type is $(l, f) = (1, 2)$ because it is has two flat faces (one on the top and one on the bottom) and a single flat segment (con-

necting the two flat faces at the points where they are furthest apart).

In Figs. 1(g)–1(j) we display a sampling of three-dimensional projections of the qutrit state space where $0 \in \{\mu, \nu, \omega\}$, so that the component s_0 is included in the projection. There are $\binom{8}{2} = 28$ of these projections. By convention, we plot the s_0 component along the vertical axis. Because the states are normalized, their projections are confined to the $s_0 = 1$ plane, and the shape is only nontrivial along the other two axes. It follows that the $s_0 = 1$ section of the projection onto the three-dimensional subspace $(0\nu\omega)$ is equivalent to the projection onto the two-dimensional subspace $(\nu\omega)$. All two-dimensional projections of the qutrit state space can be classified into different types in a manner analogous to the

three-dimensional projections, namely, by the number of flat segments in their one-dimensional boundary. We denote this number by l. For example, the projection in Fig. 1(g) is of type $l = 3$ because it has three flat segments in its boundary. Note that the projections depicted in Figs. 1(h) and 1(j) are of the same type, $l = 0$, because they both have no flat segments in their boundary. This pair of examples illustrates the fact that two projections of the same type can still have distinct shapes.

Although we do not present an exhaustive set of projection types in Fig. 1, our sampling is sufficient to give a sense of the geometric features that are characteristic of the qutrit state space.

We now turn to the projections of $\mathcal{E}^{\text{qutrit}}$. In Fig. 1 we display projections of $\mathcal{E}^{\text{qutrit}}$ alongside the corresponding projection of $\mathcal{S}^{\text{qutrit}}$. As before, therefore, we show projections where $0 \notin \{\mu, \nu, \omega\}$ appear in Figs. 1(a)–1(f), while projections where $0 \in \{\mu, \nu, \omega\}$ (i.e., which include the e_0 component) appear in Figs. 1(g)–1(j). Given that for effects, unlike states, the component e_0 can be any value in the interval $[0,1]$, even the projections where $0 \in \{\mu, \nu, \omega\}$ are nontrivial along all three axes.

It is useful to take note of some general features of the shapes appearing in Fig. 1. The set of trace-one quantum effects is always precisely the same as the set of normalized quantum states, $\text{cvxhull}(\text{AtmQEffs}(\mathbb{C}^d)) = \text{QStates}(\mathbb{C}^d)$. It follows that the $e_0 = 1/3$ section of the space of GPT effect vectors for a qutrit (which describes the set of trace-one quantum effects in the generalized Bloch representation we are considering) should be of precisely the same shape as the $s_0 = 1$ section of the space of GPT states for a qutrit. This also holds true for every three-dimensional projection where $0 \in \{\mu, \nu, \omega\}$: the $e_0 = 1/3$ section of the projected effect space should coincide with the $s_0 = 1$ section of the projected state space. This is precisely what is seen in Figs. 1(g)–1(j). Furthermore, in those three-dimensional projections where $0 \notin \{\mu, \nu, \omega\}$, we expect the projection of the effect space to be symmetric under inversion about the origin, for the reasons discussed below Eq. (12). This feature is also evident in Figs. 1(a)–1(f).

The three-dimensional projection depicted in Fig. 1(g) is particularly intuitive to interpret. It describes a classical three-level system, i.e., a classical trit. This projection is associated with the triple of Gell-Mann matrices λ_0, λ_3, and λ_8, which form a commutative algebra. It follows that the joint eigenstates of these three Gell-Mann matrices describe a triple of perfectly distinguishable states, and the associated projection of the qutrit state space is the convex hull of these, i.e., a triangle, which is the state space of a classical trit. Meanwhile, the associated projection of the qutrit effect space is the corresponding three-dimensional hypercube of classical effects [41].

Regarding the precise shapes of the other projections depicted in Fig. 1, we restrict ourselves to a few observations.

The fact that Fig. 1(a) depicts shapes that are isomorphic to a Bloch ball of states and a Bloch ball of trace-one effects follows from the fact that the algebra defined by the four operators $\{\frac{2}{3}\lambda_0 + \frac{1}{\sqrt{3}}\lambda_8, \lambda_1, \lambda_2, \lambda_3\}$ is, in fact, the Pauli algebra, and so describes a qubit embedded in the qutrit.

Meanwhile, the three operators $\{\frac{2}{3}\lambda_0 + \frac{1}{\sqrt{3}}\lambda_8, \lambda_1, \lambda_3\}$ describe the Pauli algebra without the Pauli Y, and so are

associated with a real-amplitude qubit (or "rebit"). The state space projection depicted in Fig. 1(h), therefore, is the Bloch disk associated with a rebit. The effect space projection depicted in Fig. 1(h), meanwhile, is a bit more complicated. It is the convex hull of the trace-one rebit effects (the Bloch disk at $e_0 = 1/3$), together with their complements relative to the qutrit unit effect (the trace-two rebit effects, depicted as a disk at $e_0 = 2/3$), as well as the zero effect and the qutrit unit effect.

More generally, some understanding of the shapes of the (μ, ν, ω) projections (of both the state and effect spaces) can typically be gleaned by considering the algebra defined by the operators λ_μ, λ_ν, and λ_ω and how this algebra embeds in the full qutrit algebra.

Quantum theory satisfies the no-restriction hypothesis, so the set of logically possible qutrit states is precisely equal to the set of physically realizable qutrit states, and the set of logically possible qutrit effects is precisely equal to the set of realizable qutrit effects. If the GPT describing the three-level system of our experiment were to differ from quantum theory such that it *failed* to satisfy the no-restriction hypothesis, then this would manifest as a gap between the physically realizable and logically possible state spaces (equivalently, a gap between the physically realizable and logically possible effect spaces). Any such gap would also be manifest in some of the three-dimensional projections of these.

IV. DESCRIPTION OF THE EXPERIMENT

Our experimental setup is shown in Fig. 2. A heralded single-photon source uses a continuous-wave diode laser with output at 404 nm that is focused with a lens onto a 15 mm PPKTP crystal to perform Type-II SPDC. Excess 404 nm light is filtered out with a long-pass filter (LPF), and resulting 808 nm photons are split using a polarizing beam splitter (PBS2). The transmitted photons are received by an avalanche photo diode (APD), denoted *Dh*, which heralds the arrival of the reflected photons within a coincidence window of 3 ns. To control polarization drifting effects characteristic to the laser, the reflected port of PBS1 is actively monitored with a power meter (PM100D) and is connected via a feedback loop to a motorized half wave plate (HWP) to ensure optimal crystal pump polarization at all times.

Photons from the source are coupled into a single-mode fiber with polarization control (PC) and are passed to an interferometer in a displaced Sagnac configuration. A calcite beam displacer (CBD) is used with the fiber PC to ensure vertically polarized photons are incident to the interferometer. Different states of the three-level system are prepared using motorized quarter wave plate (QWP) and HWP pairs P1 and P2 (highlighted in blue in Fig. 2). Preparation stage P1 is used to prepare a two-level state with modes $|H_1\rangle$ and $|V_1\rangle$. These modes are split on PBS3. Mode $|H_1\rangle$ travels counterclockwise around the first interferometer loop and returns to PBS3 unchanged. Mode $|V_1\rangle$ travels clockwise around the first interferometer loop and is addressed by preparation stage P2 where the third mode of the three-level state, denoted $|H_2\rangle$, is prepared. Each three-level state is fully constructed once all three modes recombine on PBS3. The measurement is varied using motorized HWP-QWP pairs M1 and M2 (highlighted
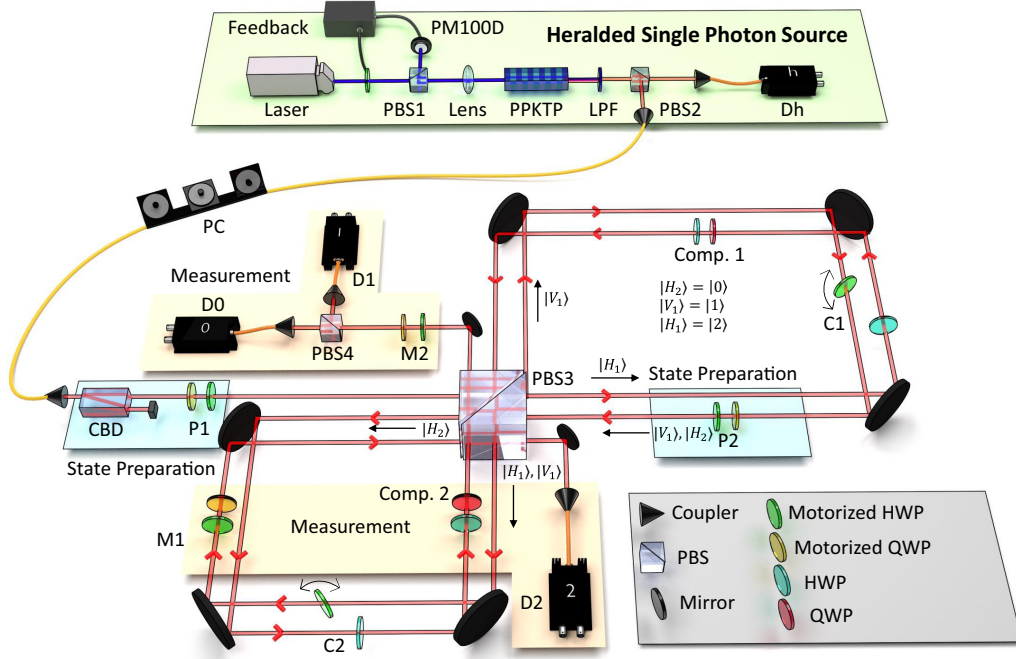
FIG. 2. Single photons at 808 nm are generated via type II SPDC using a continuous wave laser diode at 404 nm focused onto a PPKTP crystal with a lens. Polarization drifting effects of the source are actively compensated using a power meter (PM100D) connected via a feedback loop to a motorized HWP. Excess 404 nm light is filtered out with a long-pass filter (LPF). Photons detected at APD *Dh* are used to herald the arrival of photons arriving at APDs *D*0 (outcome 0), *D*1 (outcome 1), and *D*2 (outcome 2) through the displaced Sagnac interferometer with a coincidence window of 3 ns. Three-level states are prepared in the modes $|H_1\rangle$, $|V_1\rangle$, and $|H_2\rangle$ using QWP-HWP pairs P1, P2 (highlighted in blue) and are measured using HWP-QWP pairs M1, M2 (highlighted in orange). The interference visibility of each interferometer is actively compensated using two HWP pairs *C*1 and *C*2. One HWP of each pair is placed on a pivoting motorized mount, allowing for fine control of the relative optical path length. The nonmotorized wave plates, labeled as Comp. 1 and Comp. 2 and shown in blue and red, are used to balance the optical path lengths of the interferometer arms and always have optical axis oriented along the horizontal. The arrows shown within the interferometer arms are displayed for clarity of photon direction.

in orange in Fig. 2), and the photon is detected at one of three APDs *D*0, *D*1, or *D*2, corresponding to a three-outcome measurement. In our analysis, outcomes 1 and 2 are coarse grained to give a binary-outcome measurement. The nonmotorized QWP-HWP pairs in Fig. 2 (shown in blue and red and labeled Comp. 1 and Comp. 2) are used to balance the optical path lengths of the interferometer arms and always have the optical axis oriented along the horizontal.

The beams in the interferometer arms are displaced to about 5 mm. All wave plates within the interferometer arms have 3.5 mm holes allowing one beam to pass freely through. The interferometer is inside a box to reduce air currents and improve long-term stability. Relative phases of the interferometer arms are controlled using tilted HWPs placed on pivoting motors (displayed skewed in wave plate pairs C1 and C2). The interference visibility of both interferometers is automatically checked and corrected every 20 minutes as it was noted that interferometer phase began to drift after approximately 30 minutes. The compensation scheme is as follows: For a specific experimental configuration in which maximum interference should occur on PBS3 and in which the count rate at detector D1 should be minimal, the motorized HWPs in either C1 or C2 (depending on the interferometer being compensated) is pivoted until a threshold count rate is achieved at detector D1, which corresponds to an optimal

interference visibility. We obtain an optimal interference visibility for the first interferometer (top in Fig. 2) of 0.972 and an optimal interference visibility for the second interferometer (bottom in Fig. 2) of 0.984. The discrepancy between the two optimal visibility values is explained by the fact that the second interferometer has a shorter optical path length than that of the first, making it intrinsically more stable.

To choose our experimental preparation settings, we first generate a number of random qutrit quantum states that are distributed according to the SU(3) invariant Haar measure. This is accomplished by sampling a number of 3×3 unitary matrices according to the Haar distribution, whose first columns are then taken to be a set of random pure quantum states [42,43]. We then compute the HWP and QWP angles necessary to prepare each of these random states in our experimental setup and use these angles to set the QWP-HWP pairs P1 and P2 shown in Fig. 2.

Every rank-1 projective effect $Q_j = |\psi_j\rangle\langle\psi_j|$ can be matched with the state to which it responds with unit probability, namely, $\rho_j = |\psi_j\rangle\langle\psi_j|$, so that $\mathrm{Tr}[\rho_j Q_j] = \mathrm{Tr}[|\psi_j\rangle\langle\psi_j|\psi_j\rangle\langle\psi_j|] = 1$. We choose our set of measurements such that $\{Q_j, \bar{Q}_j\} = \{|\psi_j\rangle\langle\psi_j|, I - |\psi_j\rangle\langle\psi_j|\}$, where each $|\psi_j\rangle\langle\psi_j|$ corresponds to one of the sampled preparation states. Here Q_j is the effect corresponding to outcome 0 and $\bar{Q}_j$ is the effect corresponding to the coarse graining of outcomes

1 and 2. Therefore, the measurement settings are the wave plate angles that must be passed to the HWP-QWP pairs M1 and M2 shown in Fig. 2 in order to implement the n projective measurements of the form $\{Q_j, \bar{Q}_j\}$. We convert our three-outcome measurement into a binary-outcome measurement in postprocessing. During the experiment, however, we measure photon counts at all three detectors independently. There are a continuum of experimental configurations such that coarse graining outcomes 1 and 2 yields the binary-outcome measurement $\{Q_j, \bar{Q}_j\}$. Every decomposition of the rank-2 projector $\bar{Q}_j$ into a pair of rank-1 projectors yields a distinct such configuration. However, ensuring that the effect Q_j is implemented at detector D0 provides us with sufficient information to conduct our GPT analysis scheme. We therefore choose *any* set of wave plate angles that associates the effect Q_j with detector D0, without imposing any constraints on the effects associated with detectors D1 and D2.

Our choices of measurement settings produce many instances in which $p(0|P_i, M_j) = 1$ in theory. However, in practice, limitations of the experimental setup always bound the probabilities obtained away from 1, so that the uncertainties on our counts never produce supernormalized outcomes. Note that although our choices of experimental preparations and measurements are informed by quantum theory, we do not assume its correctness in any of our analysis techniques.

For our first experiment, we choose $m = 100$ Haar-distributed preparation settings and $n = 100$ corresponding measurement settings for a total of $mn = 10^4$ experimental configurations. With a laser power of 0.5 mW incident to the PPKTP crystal, we report a second order correlation function value of $g^{(2)} = 0.0034$ and detect photon coincidences at a rate of approximately 2000 counts/s. We integrate each measurement configuration for 2 s. Including the time it takes for the motorized wave plates to reposition between settings, and the time it takes to actively correct for interferometer phase drift, each run of the experiment takes approximately 16 h to complete.

V. BOOTSTRAP TOMOGRAPHY IN THE GPT FRAMEWORK

A. Inferring best-fit probabilities from experimental data

We now discuss how to estimate the probability matrix D directly from experimental data. Heralded photon counts are detected at $D0$, $D1$, and $D2$. We assume Poissonian uncertainty on the counts obtained at each detector. One can obtain estimates of the probabilities $p(0|P_i, M_j)$ by dividing the number of counts detected at $D0$ by the total number of counts obtained at all three detectors. We denote these relative frequencies by $f(0|P_i, M_j)$ and from them we construct a frequency matrix F that is an estimate of the rank-k probability matrix D. The experiment was conducted twice (back to back) using the same preparation and measurement settings. We call the first data set obtained the *training set* and the second data set obtained the *test set*. The training set, with frequency matrix F^{train}, is used to find the best-fit probability matrix D that characterizes the GPT underlying the experiment. Since the frequencies in the matrix F^{train} fluctuate away from the true probabilities, F^{train} will tend to be full rank, regardless of

the rank of the underlying matrix D. Therefore, the problem at hand is to find, for each of a set of candidate ranks, the matrix D of that rank that best fits the matrix of frequencies F^{train}. Specifically, the task can be formulated as the weighted low-rank approximation problem

$$
\begin{aligned}
\text{minimize} \quad & \chi_k^{2\ \text{train}} = \sum_{i=1}^m \sum_{j=1}^n \left(\frac{F_{ij}^{\text{train}} - D_{ij}}{\Delta F_{ij}^{\text{train}}} \right)^2, \\
\text{subject to} \quad & \text{rank}(D) = k \\
& 0 \leqslant D_{ij} \leqslant 1
\end{aligned}
\tag{14}
$$

where $\Delta F_{ij}^{\text{train}}$ is the statistical uncertainty in F_{ij}^{train}.

Equation (14) can be solved using an algorithm based on an alternating least-squares approach [24]. The column of ones corresponding to the unit effect that we include in F are added in by hand and have uncertainty 0. Because D is defined as the matrix that minimizes $\chi_k^{2\ \text{train}}$, the entries in the same column of D must also be exactly 1. (If this was not the case, then $\chi_k^{2\ \text{train}}$ would be undefined for those entries.) For each of the candidate ranks k, we compute the best-fit approximation to the experimental data, denoted D_k^{realized}. To determine which D_k^{realized} is the best approximation to the true probability matrix D, we analyze the *training error* and *testing error*. The training error for rank k is simply the $\chi_k^{2\ \text{train}}$ value obtained from Eq. (14), and it is used to determine how much a given model *underfits* the data. The testing error $\chi_k^{2\ \text{test}}$ for a data set is given by

$$
\chi_k^{2\ \text{test}} = \sum_{i=1}^m \sum_{j=1}^n \left(\frac{F_{ij}^{\text{test}} - D_{k,ij}^{\text{realized}}}{\Delta F_{ij}^{\text{test}}} \right)^2,
\tag{15}
$$

and is used to determine the *predictive power* of a given model, which can be compromised by both underfitting or *overfitting* the data. Here F^{test} is the frequency matrix corresponding to the testing data set and $\Delta F_{ij}^{\text{test}}$ is the statistical uncertainty in F_{ij}^{test}. The rank-k model that minimizes the testing error has the highest predictive power and will be the model adopted as the best fit for our data.

For the 100×100 Haar-distributed experiment, the training and testing errors for different ranks are shown in Fig. 3(a). The training error (shown in blue) decreases as the candidate rank k increases from 2 to 9. It continues to decrease as the rank increases from 9 to 12, but at a slower rate. The testing error (shown in red) also decreases as the candidate rank k increases from 2 to 9; however, unlike the training error, it begins to rise again as the rank ranges from 9 to 12. Since the rank-9 model is the one that is found to minimize the testing error, it is the one that is identified as having the maximum predictive power. Furthermore, the trends just described are precisely what one expects if the rank-9 model is the transition point between underfitting and overfitting, as we now explain.

Recall that the set of probability matrices that can be realized by models of rank $k - 1$ is a *strict subset* of what can be realized by models of rank k [the rank-$(k - 1)$ models have strictly fewer parameters than the rank-k models]. If the rank-$(k - 1)$ model underfits the data relative to the rank-k model, so that it has a higher training error and also a higher testing error, then every model with rank *less than $k - 1$* only further underfits the data relative to the rank-k model and so
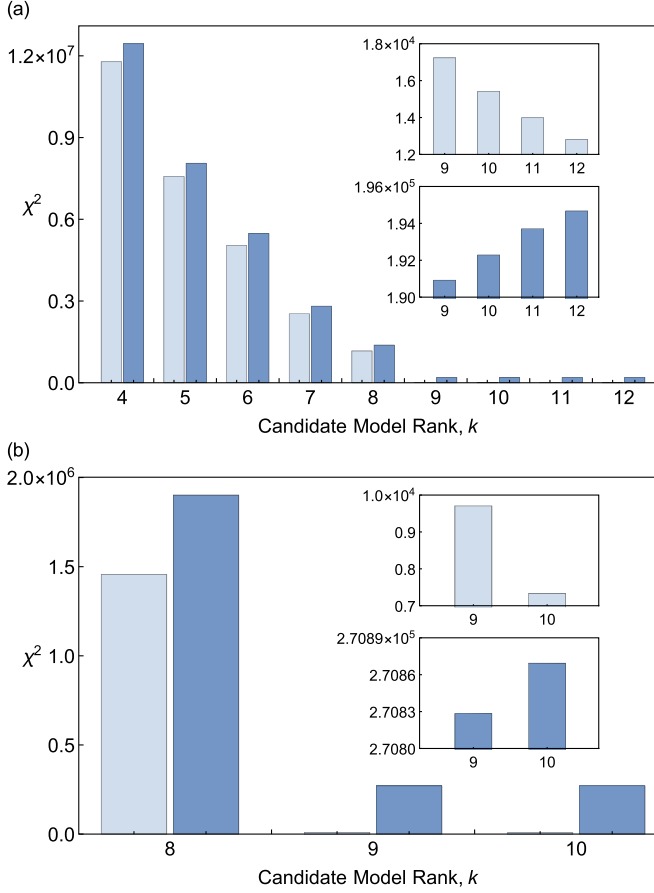
(a)

(b)

FIG. 4. Comparison between the best-fit realized probability matrices D_9^{realized} and the predicted quantum mechanical probability matrices D^{qutrit} for our two experiments. (a) A matrix plot of D_9^{realized} for the $m = n = 100$ experiment. (b) A matrix plot of D^{qutrit} corresponding to the preparation and measurement settings used in the $m = n = 100$ experiment. The mean and standard deviation of the difference between D_9^{realized} and D^{qutrit} is -0.003 and 0.03, respectively. (c) A matrix plot of D_9^{realized} for the $m = n = 415$ experiment. (d) A matrix plot of D_9^{qutrit} for the $m = n = 415$ experiment. The mean and standard deviation of the difference between D_9^{realized} and D_9^{qutrit} is -0.006 and 0.03, respectively.

FIG. 3. The training error (light blue) and testing error (dark blue) corresponding to various candidate model ranks k. (a) Training and testing errors for the $m = n = 100$ experiment. Insets show training and testing error for model ranks $k = 9$ through $k = 12$. Models with rank $k < 8$ severely underfit the experimental data as indicated by the relatively high training error. Models with ranks $k > 9$ begin to overfit the experimental data, as indicated by the increase in the testing error accompanied by a decrease in the training error. (b) Training and testing errors for the $m = n = 415$ experiment. The rank $k = 8$ model underfits the experimental data, while the rank $k = 10$ model overfits the data. For both experiments, we conclude that the best-fit GPT is rank $k = 9$ as it has the highest predictive power.

necessarily has training and testing errors that are higher than or equal to those of the rank-$(k-1)$ model. Meanwhile, if the rank-k model *overfits* the data relative to the rank-$(k-1)$ model (i.e., its greater parametric freedom causes it to fit to statistical fluctuations in the training data), so that it has a lower training error but a higher testing error, then every model with rank *greater than* k only further overfits the data relative to the rank-k model, implying that its training error can only fall further and its testing error tends to rise higher. The change in the rate of decrease of the training error after the point where the testing error is minimized is also expected. Prior to that point, an increase in rank makes a big difference for the ability of the model to fit the data. After that point, the additional parametric freedom can cause the model to fit to fluctuations in the training data, but this can yield only a slightly lower training error.
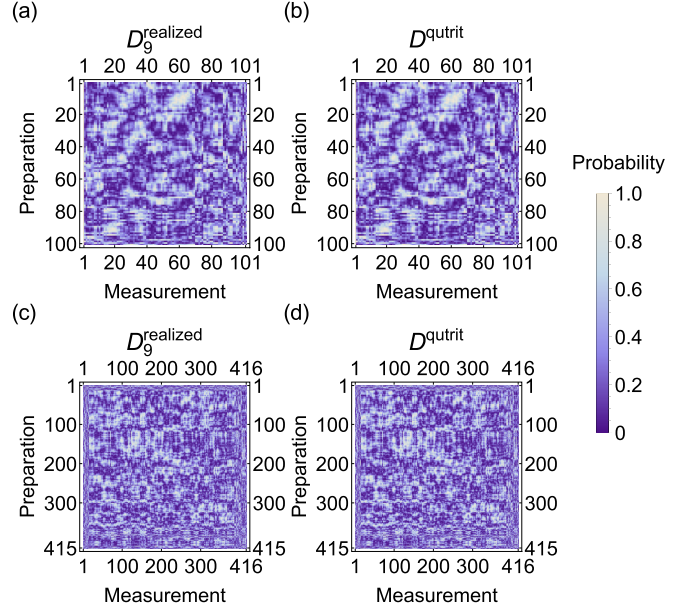
In summary, the profile of the training and test errors that we observed is precisely what one expects under the assumption of the correctness of quantum theory: the rank $k = 9$ model is the one at which there is a transition from underfitting to overfitting. The degree of underfitting monotonically increases as one decreases the rank below 9, and the degree of overfitting monotonically increases as one increases the rank above 9.

The best-fit probability matrix for our experimental data, therefore, is the rank-9 matrix D_9^{realized}. Figures 4(a) and 4(b) show a comparison between D_9^{realized} and the theoretical probability matrix predicted by qutrit quantum mechanics, denoted D^{qutrit}. Figure 4(a) shows D_9^{realized} as a matrix plot of probability values for the $m = n = 100$ experiment. The random pattern of the matrix plot is a consequence of the Haar-distributed experimental configurations. Figure 4(b) shows a similar matrix plot for D^{qutrit}. The quantum mechanical probabilities were computed using $D_{ij}^{\text{qutrit}} = \text{Tr}[\rho_i Q_j]$ with ρ_i and Q_j being the same states and effects used to compute the wave plate angles for the experimental configurations. Figures 4(a) and 4(b) show strong agreement between the experimentally realized probability matrix and the probability matrix predicted by quantum theory with the mean and standard deviation of the difference between D_{ij}^{realized} and D_{ij}^{qutrit} being -0.003 and 0.03, respectively.

B. Increasing the number of experimental configurations

The $m = n = 100$ experiment was useful for obtaining strong evidence that the dimension of the GPT underlying our experiment is nine, but the sampling of states and effects was not sufficiently dense to capture the geometries of the state and effect spaces. In this section, we outline a method for increasing the number of GPT states and effects characterized in our experiment, so that we can obtain a clearer picture of the geometries.

Simply increasing m and n and probing all state-measurement pairs leads to a large increase in the number of experimental configurations and, consequently, an increase in the experimental run time. However, to characterize a particular GPT state vector, one only needs to find its statistics on any tomographically complete subset of measurements. Similarly, to characterize any GPT effect vector, one only needs to implement it on a tomographically complete set of preparations. The cardinality of a tomographically complete set is the dimension k of our GPT vector space. It follows that one should be able to characterize m preparations and n measurements with only $k(m + n)$ experimental settings, where k is the dimension of the GPT.

Despite the success of the first experiment in identifying the dimension of the GPT describing our data, it is prudent not to assume this result when conducting our second experiment. This means that although we expect we only need nine states/effects to construct our tomographically complete sets, we use more than that to allow the new data to potentially refute the rank-9 conclusion of our first experiment. From the set of 24 normalized eigenvectors of the eight 3×3 Gell-Mann matrices (excluding identity), we choose the 15 that are distinct, denoted $\{|\psi_j\rangle\}_{j=1}^{15}$, as our tomographically overcomplete set of states. Our tomographically overcomplete set of measurements are simply the 15 binary-outcome measurements given by $\{Q_j, \bar{Q}_j\} = \{|\psi_j\rangle\langle\psi_j|, I - |\psi_j\rangle\langle\psi_i|\}$. We call the set of 15 tomographically overcomplete states (measurements) the *fiducial* set, and we use it to extend the number of states and effects we can characterize with our experiment. We sample 400 preparations and measurements according to the Haar distribution (in the same manner as our first experiment). We then pair each of the 15 fiducial states with the 400 Haar-distributed effects, and each of the 400 Haar-distributed states with the 15 fiducial effects. We also allow each fiducial state to be paired with each fiducial effect, adding 15^2 state-effect pairings. In total, the second experiment requires $15(400 + 400 + 15) = 12\,225$ state-effect pairings to be experimentally implemented, a modest 22% increase relative to the 10 000 pairings required in the first experiment, while allowing a characterization of 415 states and effects, which is more than four times the number of states and effects characterized in the first experiment. We again add the unit effect $\boldsymbol{u}$ by hand. Our data is arranged into a 415×416 frequency matrix F, where the bottom right 400×400 entries are unfilled.

We run our second experiment twice back-to-back to obtain a training and test data set and perform a similar analysis to that of our first experiment. We choose to analyze only the candidate model ranks $k \in \{8, 9, 10\}$ because our first experiment already indicated that $k = 9$ is likely to provide the best fit for our data and, as explained above, if $k = 9$ is found to be the transition point between underfitting and overfitting in the range $k \in \{8, 9, 10\}$, then one can infer that ranks $k < 8$ could only have exhibited more underfitting while ranks $k > 10$ could only have exhibited more overfitting.

We find the best-fit matrix D_k^{realized} for F^{train} using Eq. (14) and retrieve the training error for each model. For the 400×400 submatrix of unfilled entries in the frequency matrix, we take their weighting in the least-squares fit to be 0. Thus, the only constraint on the fit for the 400×400 unfilled entries is that each corresponding entry in D_k^{realized} must be a valid probability. Once D_k^{realized} is found for each candidate model, we compute the corresponding testing errors using Eq. (15). The training and testing errors for this data set were computed using only the elements of the frequency matrices that were actually measured in the experiment (i.e., the complement of the unfilled 400×400 submatrix of F^{train} and F^{test}).

The results of the training and testing error analysis for the fiducial experiment are shown in Fig. 3(b). Comparing $k = 8$ to $k = 9$, the training and testing errors clearly indicate that the $k = 8$ model underfits the experimental data relative to the $k = 9$ model, in agreement with our first experiment. Comparing $k = 10$ to $k = 9$, the fact that the training error decreases but the testing error increases indicates that the $k = 10$ model is overfitting the data relative to the $k = 9$ model. We draw the conclusion that (in agreement with our first experiment) the model with the most predictive power for our data has rank $k = 9$, and we adopt D_9^{realized} as the best-fit probability matrix.

A comparison between D_9^{realized} to the probability matrix predicted by qutrit quantum mechanics D^{qutrit} for the fiducial experiment is shown in Figs. 4(c) and 4(d). Figure 4(c) shows a matrix plot for the probabilities comprising D_9^{realized}, and Fig. 4(d) shows an analogous matrix plot for the quantum mechanically predicted probabilities. The narrow bands spanning the top and left sides of the matrix plots represent the preparations and measurements actually recorded in the experiment. These bands can be visually distinguished from the rest of the probabilities because the 15 fiducial states (effects) are more distinct from one another than the 400 states (effects) randomly sampled from the Haar distribution. The remainder of the probabilities in D_9^{realized} are filled in from the weighted least squares fit. We again observe a strong agreement between the experimentally realized probability matrix and the probability matrix predicted by quantum theory with mean and standard deviation of the difference between D_{ij}^{realized} and D_{ij}^{qutrit} of -0.006 and 0.03, respectively.

C. Constraining the shapes of the state and effect spaces

We now discuss how to decompose the $m \times n$ best-fit probability matrix D_9^{realized} into an $m \times k$ matrix S^{realized} of realized GPT state vectors and a $k \times n$ matrix E^{realized} of realized effect vectors. Recall that a decomposition $D = SE$ is not unique, because for any invertible matrix Λ, one also has $D = (S\Lambda)(\Lambda^{-1}E)$. The choice of Λ is merely a convention for how to represent the GPT state and effect spaces geometrically. We expect that because the dimension of our experimentally realized GPT is nine, we will be able to recover GPT states and effects that resemble the sets of quantum states and effects used to generate our experimental

preparations and measurements. Thus, we seek to identify the choice of Λ that yields the decomposition of D_9^{realized} that best approximates the quantum states and effects used to generate our experimental wave plate angles (specifically, the generalized Bloch representations of these states and effects). Identifying this decomposition will allow us to more easily compare our realized GPT spaces to those predicted by quantum theory.

We begin by finding an initial valid decomposition of D_9^{realized}. To obtain this initial decomposition, we follow the method outlined in Appendix D of Ref. [24]. This method is summarized as follows: We first take the *QR* decomposition of D_9^{realized}, where Q is an $m \times m$ unitary matrix, and R is an $m \times n$ upper-right triangular matrix. We then perform the singular value decomposition $QR = U \Sigma V^T$ and partition the result such that the $m \times k$ matrix of GPT states can be written as $S' = U \sqrt{\Sigma}$ and the $k \times n$ matrix of GPT effects obtained from this decomposition can be written as $E' = \sqrt{\Sigma} V^T$.

We define the matrices S^{qutrit} and E^{qutrit} to be those that consist of the Haar-distributed (and fiducial) qutrit state and effect vectors used to generate the wave plate angles for the second experiment. Given the initial decomposition $D_9^{\text{realized}} = S'E'$, the goal is to find an invertible $k \times k$ matrix Λ such that $S'\Lambda$ is as close as possible to S^{qutrit}. Therefore, we wish to find the matrix Λ that minimizes the cost function

$$\text{minimize} \quad \chi^2 = \sum_i \sum_\alpha [S_{i\alpha}^{\text{qutrit}} - (S'\Lambda)_{i\alpha}]^2,$$

$$\text{subject to} \quad \Lambda \quad \text{invertible} \tag{16}$$

where $i \in \{1, 2, \ldots, 415\}$ and $\alpha \in \{0, \ldots, 8\}$.

If $\Lambda_\star$ is the optimal transformation matrix computed from Eq. (16), we take it as our conventional choice of Λ in the decomposition $D_9^{\text{realized}} = (S'\Lambda)(\Lambda^{-1}E')$ and thereby take $S^{\text{realized}} = S'\Lambda_\star$ and $E^{\text{realized}} = \Lambda_\star^{-1}E'$ to be the matrices defining our experimentally realized GPT state and effect vectors. This choice facilitates the project of looking for potential deviations from quantum theory.

The *realized* GPT state space, denoted $\mathcal{S}^{\text{realized}}$, describes the states that were in fact realized by the experiment. The convex hull of the rows of the matrix S^{realized} yields our best estimate of $\mathcal{S}^{\text{realized}}$. It is merely an estimate due to the fact that S^{realized} is based on the relative frequencies in a finite run, rather than from long-run probabilities. Similarly, the *realized* GPT effect space, denoted $\mathcal{E}^{\text{realized}}$, describes the effects that were in fact realized by the experiment. The procedure by which we obtain our best estimate of $\mathcal{E}^{\text{realized}}$ is as follows. We begin by taking the set of GPT effect vectors defined by the columns of E^{realized}. Next, we find the set of complements of these, i.e., the set of vectors $\mathbf{u} - \mathbf{e}$ for each $\mathbf{e}$ corresponding to a column of E^{realized}. Finally, we identify the convex hull of these two sets. Note that because the unit effect vector $\mathbf{u}$ is included as the first column of E^{realized}, it follows that the zero effect vector $\mathbf{0}$ is included in the set of complements of effect vectors, and so the unit and zero effect vectors are necessarily included in the convex hull.

The set of GPT effect vectors that are logically consistent with the realized GPT state vectors, termed the *consistent* effect space and denoted $\mathcal{E}^{\text{consistent}}$, is the set of all vectors $\mathbf{e}$ such that $\forall s \in \mathcal{S}^{\text{realized}}: 0 \leqslant s \cdot e \leqslant 1$. For simplicity, we write

$\mathcal{E}^{\text{consistent}} = \text{dual}(\mathcal{S}^{\text{realized}})$. Analogously, the set of GPT state vectors that are logically consistent with the realized GPT effect vectors, termed the *consistent* state space and denoted $\mathcal{S}^{\text{consistent}}$, is the set of all vectors s such that $\forall e \in \mathcal{E}^{\text{realized}}: 0 \leqslant s \cdot e \leqslant 1$ and $s \cdot u = 1$. We write $\mathcal{S}^{\text{consistent}} = \text{dual}(\mathcal{E}^{\text{realized}})$.

We now explain how the realized and consistent state (effect) space obtained from our experiment relates to the true state (effect) space governing nature. First, the state and effects that are realized experimentally necessarily involve additional noise relative to what is possible in the true GPT. As such, we expect all of the states and effects we realize to be slightly noisy versions of the ones that are stipulated to be possible by the true GPT. Geometrically, this means that the realized state and effect vectors should be slightly *contracted*, in directions orthogonal to the normalization axis, relative to their noiseless counterparts. And this in turn implies that the state and effect vectors that are *logically consistent* with the ones we realized should be slightly *elongated* in these directions relative to their noiseless counterparts.

Second, the fact that one can only realize a *finite* number of states and effects in the experiment implies that the realized state and effect spaces will be convex polytopes even if the true state and effect spaces are not. The same is true of the logically consistent spaces that one obtains from the realized spaces. The extent to which the realized and consistent spaces approximate the true and logical spaces depends on how densely sampled the full set of states and effects is.

The experimental limitations described above imply that the realized spaces are expected to be *strictly contained* within the true spaces, $\mathcal{S}^{\text{realized}} \subset \mathcal{S}$ and $\mathcal{E}^{\text{realized}} \subset \mathcal{E}$, while the consistent spaces strictly contain the logical spaces, $\mathcal{S}^{\text{logical}} \subset \mathcal{S}^{\text{consistent}}$ and $\mathcal{E}^{\text{logical}} \subset \mathcal{E}^{\text{consistent}}$. It follows that the true GPT state (effect) space governing our three-level system must lie somewhere between the realized and consistent GPT state (effect) space inferred from our experiment. More precisely, $\mathcal{S}$ must satisfy $\mathcal{S}^{\text{realized}} \subseteq \mathcal{S} \subseteq \mathcal{S}^{\text{consistent}}$, which is equivalent to the constraint that $\mathcal{E}$ must be such that $\mathcal{E}^{\text{realized}} \subseteq \mathcal{E} \subseteq \mathcal{E}^{\text{consistent}}$. [Recall also that $\mathcal{S}$ and $\mathcal{E}$ are not independently specifiable. For instance, once $\mathcal{S}$ is specified, the possibilities for $\mathcal{E}$ are constrained to satisfy $\mathcal{E} \subseteq \text{dual}(\mathcal{S})$.] Hence, one does not *uniquely* identify the true state and effect spaces; one only delimits the scope of possibilities for what they might be.

Finally, to plot three-dimensional projections of $\mathcal{S}^{\text{realized}}$ and $\mathcal{E}^{\text{realized}}$, we rely on the fact that these convex sets are polytopes. We first identify the vectors describing the vertices of these polytopes. We then compute the projections of each of these vertices. Finally, we plot the convex hull of the resulting sub-vectors corresponding to each projection. The definitions of $\mathcal{S}^{\text{consistent}}$ and $\mathcal{E}^{\text{consistent}}$ provided above are known as the *inequality representations*. However, in order to plot projections of $\mathcal{S}^{\text{consistent}}$ and $\mathcal{E}^{\text{consistent}}$ in a similar manner to how we plot projections of $\mathcal{S}^{\text{realized}}$ and $\mathcal{E}^{\text{realized}}$, we require the so-called *vertex representations* of the spaces, which lists all the vertices that specify their convex hulls. Converting from an inequality representation to a vertex representation is highly nontrivial and requires solving the *vertex enumeration problem*. Fortunately, the solution to the problem can be computed using an algorithm first developed by Avis and Fukada [44]. Calculation of the vertex representations of our consistent spaces was performed using the LRS package in C

that implements a modified version of the vertex enumeration algorithm [45].

We end this section with some comments on how experimental imperfections manifest themselves in this data analysis technique. We begin by considering this question under the assumption of the correctness of quantum theory. Although an experimentalist may aim to achieve the ideal of *pure* states and *projective* measurements, what is realized experimentally always deviates from this ideal. It is a critical feature of bootstrap GPT tomography that the rows of the matrix S^{realized} and the columns of the matrix E^{realized} (which ultimately determine $\mathcal{S}^{\text{realized}}$ and $\mathcal{E}^{\text{realized}}$) are characterizations of the preparations and measurements that were *actually realized* in the experiment, and not characterizations of the idealized preparations and measurements that the experimentalist was targeting. It is useful to consider two distinct ways in which a realized procedure might deviate from the ideal that was being targeted: it might differ in a systematic and consistent way from the target (i.e., inaccuracy), or it might exhibit statistical fluctuations away from the target (i.e., imprecision). In our experiment, an example of *inaccuracy* is if the wave plate angles that are implemented for a given preparation procedure differ in a systematic and consistent way (across all repetitions of the procedure) from those that would achieve the particular pure state being targeted. In this case, our tomography scheme would return a best-fit GPT state that is rotated in Bloch space (i.e., the qutrit analog of the Bloch ball) relative to the target. An example of *imprecision* is if, in the ensemble of repetitions of a given preparation procedure, a wave plate angle exhibits statistical fluctuations away from the targeted value. In this case, our tomography scheme would return a best-fit GPT state that was a noisy version of the target, for instance, one that is *contracted* (in Bloch space) relative to the pure state being targeted. If one is willing to make strong assumptions about the nature of the imperfections in one's experiment, then one might hope to process the raw data to compensate for these and thereby achieve a characterization of the ideal procedures that one was targeting. (Background subtraction is an example of such a processing.) Bootstrap GPT tomography, however, simply analyzes the raw data, rather than a processed version thereof, and allows one to *infer* the nature of the imperfections. Making such inferences about experimental imperfections, however, is only possible relative to some assumption about what the correct operational theory is. This is because it is only under such an assumption that an experimentalist can have *any expectation* about what GPT states and effects will characterize some experimental procedure (even an idealized one). For example, under the assumption of the correctness of *quantum theory*, one expects certain idealized preparation procedures to be on the surface of the state space dictated by quantum theory (i.e., Bloch space), such that rotations and contractions of these GPT states relative to one's expectations can be used to infer inaccuracies and imprecisions in the actually realized procedures relative to their idealized counterparts. Given that the purpose of this article is to find a way to analyze the experimental data *without* making assumptions about which theory is correct, it is imperative *not* to make any assumptions about what GPT states describe the experimental procedures one is implementing (or putative idealized versions thereof).

It follows that the data analysis scheme one uses *must* be such that it returns a characterization of the actually realized procedures, as bootstrap GPT tomography does. To put it another way: in a theory-agnostic tomography scheme, it is not possible to imagine a processing of the raw data that can compensate for imperfections relative to some putative ideal because one is not allowed to make assumptions about the nature of this ideal.

VI. RESULTS

In Fig. 5 we present various three-dimensional projections of the realized and consistent spaces we found in our experiment. Specifically, we present the same projections that are showcased in Fig. 1 for ease of comparison with the latter. For a given three-dimensional subspace, we plot the projection of S^{realized} into this subspace as a yellow polytope, and the projection of E^{realized} as a cyan polytope. The corresponding projections of the spaces $\mathcal{S}^{\text{consistent}}$ and $\mathcal{E}^{\text{consistent}}$ are plotted as mesh polytopes alongside the projections of $\mathcal{S}^{\text{realized}}$ and $\mathcal{E}^{\text{realized}}$. For the state space projections that include the normalization component, depicted in Figs. 5(g)–5(j), the consistent state spaces are plotted in dark orange rather than as a mesh.

A visual comparison of Fig. 5 to Fig. 1 reveals that our results do conform broadly with the hypothesis that the qutrit state (effect) space lies between the realized and consistent state (effect) spaces obtained in our experiment. Strictly speaking, consistency holds if some *invertible linear transformation* of the qutrit space lies between the realized and consistent space. We shall return to the question of consistency with quantum theory at the end of this section.

The question of whether the no-restriction hypothesis holds for the true GPT governing nature is the question of whether all logically possible states (effects) are included in the state (effect) space. In other words, it is the question of whether there is a gap between $\mathcal{S}$ and $\mathcal{S}^{\text{logical}}$ or between $\mathcal{E}$ and $\mathcal{E}^{\text{logical}}$. As noted above, however, experimental noise and finite sampling imply that the gap between the realized and consistent spaces is expected to be strictly larger than the gap between the true and logical spaces. Seeing a nontrivial gap of the former type, therefore, does not imply that a gap of the latter type exists. Nonetheless, one can put an *upper bound* on possible violations of the no-restriction hypothesis.

The gap between $\mathcal{S}^{\text{realized}}$ and $\mathcal{S}^{\text{consistent}}$ therefore delimits the scope of possible deviations from quantum theory. In particular, they bound how far any candidate GPT describing our three-level system may deviate from the no-restriction hypothesis. We will here focus on bounding the extent to which the no-restriction hypothesis might fail based on the gap between $\mathcal{S}^{\text{realized}}$ and $\mathcal{S}^{\text{consistent}}$. We turn now to the question of how to quantify this gap.

In the investigation of two-level systems undertaken in Ref. [24], the gap was quantified by the ratio of the *volumes* of $\mathcal{S}^{\text{realized}}$ and $\mathcal{S}^{\text{consistent}}$. This approach is not feasible here—the fact that the state spaces are nine-dimensional and the fact that the consistent state space has a very large number of vertices made it intractable to compute the volume ratio in a straightforward manner. We have opted, therefore, to consider instead a ratio of *linear dimensions*, specifically, the distances
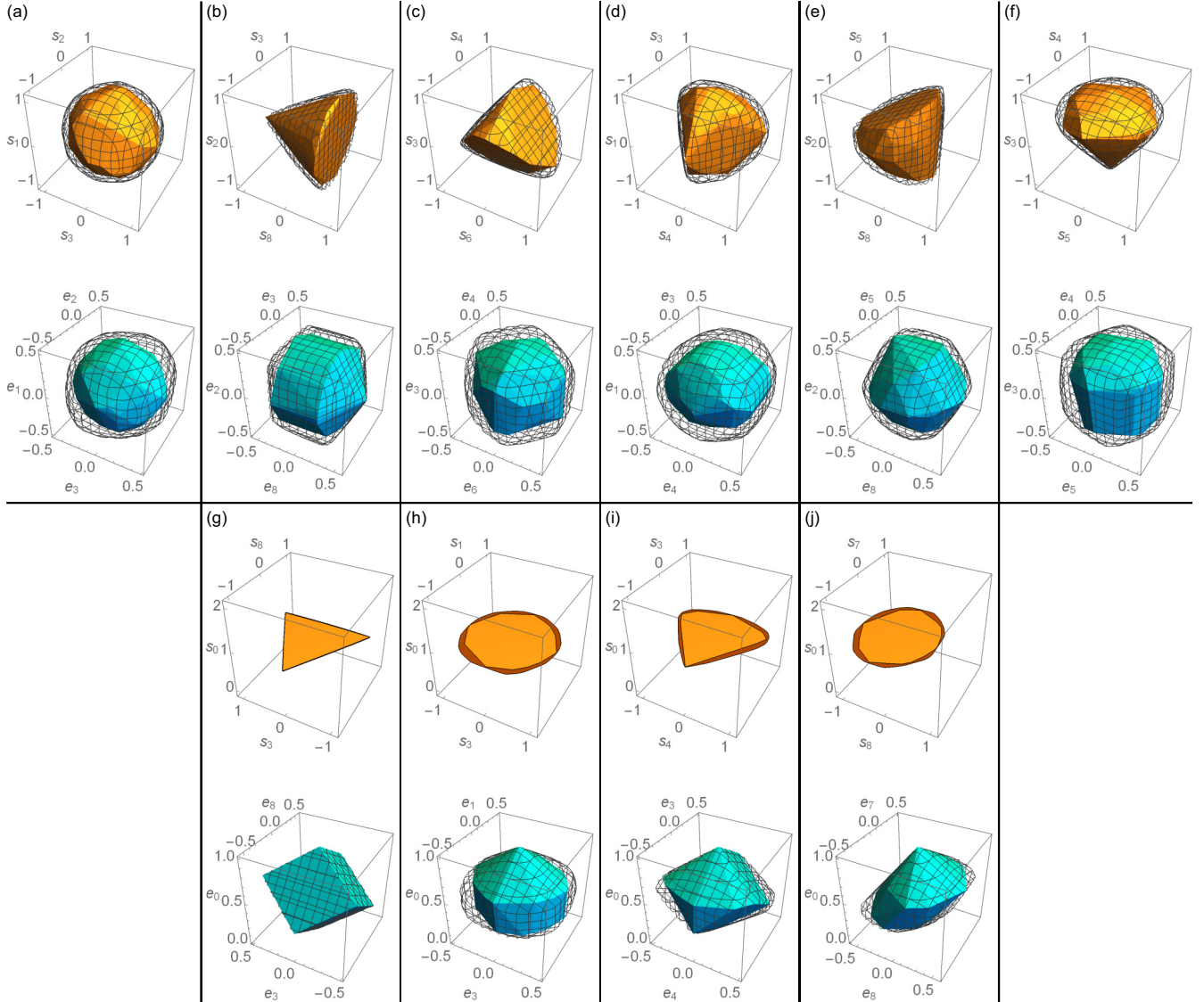
FIG. 5. A sampling of three-dimensional projections of the realized GPT state space $\mathcal{S}^{\text{realized}}$ (top polytope of each subfigure, displayed in yellow) and of the realized GPT effect space $\mathcal{E}^{\text{realized}}$ (bottom polytope of each subfigure, displayed in cyan). The mesh polytopes surrounding each of the yellow polytopes depict three-dimensional projections of the consistent GPT state space $\mathcal{S}^{\text{consistent}}$, while the mesh polytopes surrounding each of the cyan polytopes depict three-dimensional projections of the consistent GPT effect space $\mathcal{E}^{\text{consistent}}$. The true state and effect spaces of the GPT describing nature must lie somewhere between the realized and consistent GPT spaces. The specific projections depicted in (a)–(i) are the same as in Fig. 1.

from the "center" of the space of normalized states, the vector $(1, 0, \ldots, 0)$, to the surface of the realized and consistent state spaces along some direction. As this ratio of distances will vary with the direction being considered, we quantify the gap by averaging the ratio over a large sampling of random directions. We consider directions defined by the generalized Bloch vectors of pure quantum states, and we sample these according to the Haar measure. For each ray, we divide the norm of the vector extending to the intersection point with $\mathcal{S}^{\text{realized}}$ by the norm of the vector extending to the intersection point with $\mathcal{S}^{\text{consistent}}$ to obtain our *linear dimension ratio*. For a set of 1000 rays, the average linear dimension ratio between $\mathcal{S}^{\text{realized}}$ and $\mathcal{S}^{\text{consistent}}$ is computed to be 0.80, with an associated standard deviation of 0.04. Any candidate GPT

describing our three-level system that admits an average linear dimension ratio much smaller than 0.80 exhibits a greater failure of the no-restriction hypothesis than is consistent with our experimental findings.

Note that the average linear dimension ratio within each of the various *three-dimensional projections* of the realized and consistent spaces is not related in a simple manner to the average linear dimension ratio for the full-dimensional spaces. This is because projections need not correspond to sections that pass through the origin. Nonetheless, a visual inspection of Fig. 5 reveals that the average linear dimension ratio for many of these projections might well be close to 0.80.

To conclude this section, we return to the question of whether an invertible linear transformation of the qutrit state

space $\mathcal{S}^{\text{qutrit}}$ fits between the realized and consistent state spaces, $\mathcal{S}^{\text{realized}}$ and $\mathcal{S}^{\text{consistent}}$. The question turns out to be more nuanced than it appears at first glance. What complicates matters is that the rows of the matrix S^{realized} are, strictly speaking, only finite-sample *estimates* of the states in $\mathcal{S}^{\text{realized}}$. Recall that, by definition, a given state vector in $\mathcal{S}^{\text{realized}}$ specifies the long-run probabilities that would be obtained by the associated realized preparation procedure, and finite-run relative frequencies need not coincide with these long-run probabilities. Finite-run estimates of the realized state vectors will, therefore, exhibit fluctuations away from the realized state vectors. In particular, the finite-run estimates of certain realized state vectors might be anomalously long. Similarly, the finite-run estimates of certain realized *effect* vectors might be anomalously long, implying that the finite-run estimates of certain *consistent* state vectors might be anomalously *short*. Together these two anomalies can lead to a situation wherein $\mathcal{S}^{\text{qutrit}}$ does not in fact fit between the estimate of $\mathcal{S}^{\text{realized}}$ and the estimate of $\mathcal{S}^{\text{consistent}}$.

Given that the geometry of the qutrit state space is nontrivial, and given that our estimate of the consistent state space has a very large number of vertices, it is not straightforward to solve this decision problem. We therefore resort to a heuristic assessment of the question. We again consider the 1000 rays sampled according to the Haar measure from the directions defined by the generalized Bloch vectors of pure qutrit quantum states. Let l be an index that runs over these rays. For each l, define s_l^{realized} to be the vector in the direction of the lth ray that lies on the boundary of $\mathcal{S}^{\text{realized}}$, and define $s_l^{\text{consistent}}$ to be the one that lies on the boundary of $\mathcal{S}^{\text{consistent}}$. Because all pure qutrit states have generalized Bloch vectors of the same norm, if the longest of the s_l^{realized} is shorter than the shortest of the $s_l^{\text{consistent}}$, i.e., if $\max_l ||s_l^{\text{realized}}|| \leqslant \min_l ||s_l^{\text{consistent}}||$, then there is a linear transformation of $\mathcal{S}^{\text{qutrit}}$ (in particular, one that preserves the relative norm of pure states) which is straddled by our estimate of $\mathcal{S}^{\text{realized}}$ and our estimate of $\mathcal{S}^{\text{consistent}}$ at least along the 1000 random directions. We find $\max_l ||s_l^{\text{realized}}|| = 1.11$ and $\min_l ||s_l^{\text{consistent}}|| = 1.13$, so the straddling condition does hold for the 1000 random directions. In summary, our experimental results seem to be consistent with the hypothesis that quantum theory is the true GPT underlying our experiment.

VII. CONCLUSIONS

In this paper, we reconstructed the GPT state and effect spaces for a three-level photonic system directly from experimental data. The tomographic scheme utilized is bootstrap in that it does not require any prior characterizations of the preparations and measurements used in the experiment. Furthermore, neither our experimental scheme nor our analysis assumed the correctness of quantum theory. From the statistics obtained in our experiment, we were able to infer that if our sets of states and effects are tomographically complete, then the dimension of the underlying GPT for our system is nine. This inference was accomplished by computing the training and testing error [Eqs. (14) and (15)] for the best-fit probability matrices of different ranks, and identifying the model that tests best. Although our experiment provided an oppor-

tunity to discover that the dimension of the GPT governing our three-level system differed from the quantum mechanical predictions, our analysis did not yield any evidence in favour of such a deviation.

We have also identified the scope of possible shapes of the true GPT state and effect spaces, assuming they are indeed nine-dimensional. The true GPT state space must lie between the realized GPT state space $\mathcal{S}^{\text{realized}}$ and the consistent GPT state space $\mathcal{S}^{\text{consistent}}$. Similarly, the true GPT effect space must lie between the realized GPT effect space $\mathcal{E}^{\text{realized}}$ and the consistent GPT effect space $\mathcal{E}^{\text{consistent}}$. We displayed the geometries of our estimates of the realized and consistent state and effect spaces as three-dimensional projections of the full nine-dimensional spaces in Fig. 5. In addition, we were able to place quantitative bounds on how much any proposed alternative to quantum theory may violate the no-restriction hypothesis by identifying the average ratio of linear dimensions of $\mathcal{S}^{\text{realized}}$ and of $\mathcal{S}^{\text{consistent}}$ to be 0.80. Even though our experimental data provides some room for violations of the no-restriction hypothesis, it is notable that we observe good agreement in the shapes of our experimental GPT spaces when comparing them to the quantum mechanical predictions.

The results obtained in this paper provide constraints on the scope of possible deviations from quantum theory. There are many avenues to explore in conducting further experiments of this kind. The obvious extension of this work would be to undertake a GPT characterization of higher-dimensional systems or of composite systems. For example, one might undertake a GPT characterization of a pair of qubits and test the validity of principles that concern such systems, such as the principle of local tomography. Another interesting possible extension of this work is to undertake a GPT characterization of channels by performing the GPT equivalent of quantum process tomography.

The results of our analysis and those of Ref. [24] demonstrate the utility of the GPT framework in adjudicating between quantum theory and its alternatives in the landscape of possible physical theories.

ACKNOWLEDGMENTS

[1] L. Masanes and M. P. Müller, New J. Phys. **13**, 063001 (2011).

[2] A. B. Sainz, Y. Guryanova, A. Acín, and M. Navascués, Phys. Rev. Lett. **120**, 200402 (2018).

[3] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **84**, 012311 (2011).

[4] G. C. Ghirardi, A. Rimini, and T. Weber, Phys. Rev. D **34**, 470 (1986).

[5] I. C. Percival, Proc. R. Soc. Lond. A **447**, 189 (1994).

[6] P. C. E. Stamp, Phil. Trans. R. Soc. A **370**, 4429 (2012).

[7] G. J. Milburn, Phys. Rev. A **44**, 5401 (1991).

[8] M. Navascués, Y. Guryanova, M. Hoban, and A. Acín, Nat. Commun. **6**, 6288 (2015).

[9] R. D. Sorkin, Mod. Phys. Lett. A **09**, 3119 (1994).

[10] U. Sinha, C. Couteau, T. Jennewein, R. Laflamme, and G. Weihs, Science **329**, 418 (2010).

[11] D. K. Park, O. Moussa, and R. Laflamme, New J. Phys. **14**, 113025 (2012).

[12] T. Kauten, R. Keil, T. Kaufmann, B. Pressl, Č. Brukner, and G. Weihs, New J. Phys. **19**, 033017 (2017).

[13] A. Peres, Phys. Rev. Lett. **42**, 683 (1979).

[14] S. L. Adler, Nucl. Phys. B **415**, 195 (1994).

[15] L. Hardy, in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, edited by H. Halvorson (Cambridge University Press, Cambridge, 2009), Chap. 11, pp. 409–442.

[16] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).

[17] J. Barrett, Phys. Rev. A **75**, 032304 (2007).

[18] L. Hardy, arXiv:quant-ph/0101012.

[19] C. M. Scandolo, arXiv:1901.08054.

[20] M. Weilenmann and R. Colbeck, Quantum **4**, 236 (2020).

[21] G. M. D'Ariano, Probabilistic theories: What is special about quantum mechanics? in *Philosophy of Quantum Information and Entanglement*, edited by A. Bokulich and G. Jaeger (Cambridge University Press, Cambridge, 2010), pp. 85–126.

[22] A. J. Short and J. Barrett, New J. Phys. **12**, 033034 (2010).

[23] P. Janotta and H. Hinrichsen, J. Phys. A: Math. Theor. **47**, 323001 (2014).

[24] M. D. Mazurek, M. F. Pusey, K. J. Resch, and R. W. Spekkens, PRX Quantum **2**, 020302 (2021).

[25] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, Phys. Rev. A **64**, 052312 (2001).

[26] J. S. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. L. Pregnell, Ch. Silberhorn, T. C. Ralph, J. Eisert, M. B. Plenio and I. A. Walmsley, Nat. Phys. **5**, 27 (2009).

[27] C. Eltschka, M. Huber, S. Morelli, and J. Siewert, Quantum **5**, 485 (2021).

[28] For instance, Ref. [46] proposes a modification to quantum theory that does not impact qubits, but that alters the predictions for qutrits.

[29] J. Xie, A. Zhang, N. Cao, H. Xu, K. Zheng, Y.-T. Poon, N.-S. Sze, P. Xu, B. Zeng, and L. Zhang, Phys. Rev. Lett. **125**, 150401 (2020).

[30] P. J. Daley, K. J. Resch, and R. W. Spekkens, arXiv:2108.00053 [quant-ph].

[31] G. Schwarz, Ann. Stat. **6**, 461 (1978).

[32] S. Geman and C.-R. Hwang, Ann. Stat. **10**, 401 (1982).

[33] S. K. Goyal, B. N. Simon, R. Singh, and S. Simon, J. Phys. A: Math. Theor. **49**, 165203 (2016).

[34] G. Kimura, Phys. Lett. A **314**, 339 (2003).

[35] E. Brüning, H. Mäkelä, A. Messina, and F. Petruccione, J. Mod. Opt. **59**, 1 (2012).

[36] I. Bengtsson, S. Weis, and K. Życzkowski, in *Geometric Methods in Physics*, edited by P. Kielanowski, S. T. Ali, A. Odzijewicz, M. Schlichenmaier, and T. Voronov (Springer Basel, Basel, 2013), pp. 175–197.

[37] I. Bengtsson and K. Życzkowski, The space of density matrices. in *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2006), pp. 209–232.

[38] K. Życzkowski, https://pirsa.org/16120010 (2016).

[39] K. Szymański, S. Weis, and K. Życzkowski, Linear Algebra Appl. **545**, 148 (2018).

[40] We draw the set of random qutrit quantum states and effects from the SU(3) invariant Haar distribution. We discuss this method in Sec. IV.

[41] This is the qutrit analog of how the projection of the space of *qubit* states into any two-dimensional subspace which includes the identity yields a line segment state space, which is the state space of a classical bit, while the corresponding projection of the space of qubit effects yields a diamond-shaped effect space, which is the effect space of a classical bit.

[42] J. Emerson, Y. S. Weinstein, M. Saraceno, S. Lloyd, and D. G. Cory, Science **302**, 2098 (2003).

[43] F. Mezzadri, Notices Amer. Math. Soc. **54**, 592 (2007).

[44] D. Avis and K. Fukuda, Discrete Comput. Geom. **8**, 295 (1992).

[45] D. Avis, lrs home page, http://cgm.cs.mcgill.ca/~avis/C/lrs.html (2021).

[46] T. Rudolph, arXiv:quant-ph/0608120.