Data-driven inference, reconstruction, and observational completeness of quantum devices

Michele Dall'Arno,[1,2,*] Francesco Buscemi,[3,†] Alessandro Bisio,[4,5] and Alessandro Tosini[4,5]

[1]*Yukawa Institute for Theoretical Physics, Kyoto University, Kitashirakawa Oiwakecho, Sakyoku, Kyoto 606-8502, Japan*
[2]*Faculty of Education and Integrated Arts and Sciences, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan*
[3]*Graduate School of Informatics, Nagoya University, Chikusa-ku, 464-8601 Nagoya, Japan*
[4]*Quit group, Dipartimento di Fisica, Università degli studi di Pavia, via Bassi 6, 27100 Pavia, Italy*
[5]*Istituto Nazionale di Fisica Nucleare, Gruppo IV, via Bassi 6, 27100 Pavia, Italy*

The range of a quantum measurement is the set of outcome probability distributions that can be produced by varying the input state. We introduce data-driven inference as a protocol that, given a set of experimental data as a collection of outcome distributions, infers the quantum measurement which is (i) consistent with the data, in the sense that its range contains all the distributions observed, and (ii) maximally noncommittal, in the sense that its range is of minimum volume in the space of outcome distributions. We show that data-driven inference is able to return a unique measurement for any data set if and only if the inference adopts a (hyper)spherical state space (for example, the classical or the quantum bit). In analogy to informational completeness for quantum tomography, we define observational completeness as the property of any set of states that, when fed into any given measurement, produces a set of outcome distributions allowing for the correct reconstruction of the measurement via data-driven inference. We show that observational completeness is strictly stronger than informational completeness, in the sense that not all informationally complete sets are also observationally complete. Moreover, we show that for systems with a (hyper)spherical state space, the only observationally complete simplex is the regular one, namely, the symmetric informationally complete set.

I. INTRODUCTION

In quantum theory, as a consequence of the Born rule, a measurement can always be seen as a linear mapping from the set of states (i.e., density operators) into the set of probability distributions over the measurement outcomes.[1] In fact, some axiomatic approaches *identify* quantum measurements with the set of such mappings: in such a case, the resulting distribution, i.e., the image of the state of the system undergoing the measurement, receives the natural operational interpretation of distribution over the measurement outcomes [1].

When thinking of measurements as linear mappings, the image of the set of *all* states under a given measurement—also known as the measurement's range—turns out to be a very important mathematical object in quantum measurement theory. For example, given two quantum measurements, the range of one includes the range of the other, if and only if the former can simulate the latter by means of a suitable statistical transformation [2–4], independently of the state being measured. Quantum measurements, hence, can be compared by comparing the corresponding ranges, thus establishing a deep connection between quantum measurement theory and the theory of majorization and statistical comparison [5,6], with ramified consequences in both theory and applications [7–16].

In this paper we exploit the correspondence between measurements and their ranges to propose a method to derive an inference about an unknown quantum measurement, based solely on the outcome distributions observed, without any knowledge about the exact states that gave rise to such distributions. As we observe in what follows, such a method can be naturally divided into two parts. In the first part, one defines an inference rule, which formulates in an abstract way the rules that we choose to use when reasoning in the presence of incomplete information. For the problem at hand, such rules accept as input a set of outcome distributions and return as output a set of quantum measurements. For this reason, we name our inference rule "data-driven inference (DDI) of quantum measurements." The measurements inferred via DDI are consistent with the input data and are maximally noncommittal, in the sense that their ranges contain the input data and are of minimum volume in the space of outcome distributions. DDI need not aim to infer the "true" quantum measurement, as there need not be any such "entity" at this stage. As an application we implement an algorithm for the machine learning of qubit measurements based on data-driven inference, and we test it on data generated by the IBM Q Experience quantum computer.

---

*dallarno.michele@yukawa.kyoto-u.ac.jp
†buscemi@i.nagoya-u.ac.jp

[1]Since the Born rule is, in fact, bilinear in the state-measurement pair, also the opposite is true, namely, that any state induces a linear mapping from measurements into probability distributions. For this reason, in the Appendix A the formalism is developed for both states and measurements. However, for the sake of concreteness, the narrative in the main text mostly follows the task of measurement inference.

In the second part, one needs to show that it is indeed possible to construct a real experiment so that DDI leads to the correct assignment for the unknown measurement. The goal here is reminiscent of that of conventional quantum measurement tomography [17–23], namely, the reconstruction of an unknown measurement from the statistics collected in a sequence of experimental trials. However, while the data analysis performed in measurement tomography requires the knowledge of the states that were fed into the unknown measurement, DDI reconstruction only requires the analysis of the bare outcome distributions: the state preparator could, for example, emit a different unknown state at each repetition of the experiment, and DDI reconstruction would still be applicable.[2]

In what follows, we expound the theory of data-driven inference and reconstruction for finite dimensional systems. As this is based on the correspondence between measurements and their ranges, three main problems arise and are addressed.

The first problem is to seek for a general method to infer a range given a set of outcome distributions. As a possible solution, in what follows, we propose that the measurement range to be inferred, in the face of a set of experimental data, should be the *smallest one containing all the observed data*. Recalling that the range of a measurement is directly related to the ability to simulate other measurements [2], our principle is equivalent to say that the measurements to be inferred should be the weakest possible, compatibly with the data. Our inference rule hence encapsulates a principle of "self-consistent minimality" that we believe constitutes a natural way to reason in the presence of incomplete information. We show that the *only* systems for which DDI always leads to a unique range for any set of data, among all generalized probabilistic theories [24–30], are those with (hyper)spherical state space, such as the classical and the quantum bit. This can be interpreted as an "epistemic reconstruction" of such systems, regarded as epistemic hypotheses onto which to base our reasoning, rather than actual entities to be operationally characterized [31–34].

The second problem consists of understanding to which extent the correspondence between a measurement and its range can be inverted, that is, to what extent a measurement can be characterized if only its range is given. In this respect, in what follows, we show that the correspondence measurement range is invertible, but only up to the action of a symmetry transformation leaving the state space of the system invariant. This is something to be expected when directly working in the space of outcome distributions, and we consider this to be a feature, rather than a limitation, of DDI.

The third problem is to understand how an experimentalist, in complete control of their laboratory, can produce experimental data, which are rich enough to reconstruct, via DDI, the "correct" range of a measurement. That is, we want to understand whether, in order to recover the correct range by DDI, an infinite set of states needs to be prepared and sent through the measurement apparatus, or whether a finite set of states, and possibly the same ones for any measurements, suffice (we remark that, in contrast to the data analysis of quantum measurement tomography, DDI does not require the knowledge of such states). This problem is analogous to the problem in quantum tomography to construct a standard apparatus that work whatever it is to be reconstructed. As the problem in quantum tomography is solved by an informationally complete apparatus, the analogous problem in DDI reconstruction is solved by what we call an *observationally complete* (OC) apparatus. More precisely, OC sets of states are sets whose image contains the same statistical information as the entire range. Of course, since DDI does not rely on the knowledge of the states, it is not possible for DDI to certify whether the states fed into the measurement were OC.

We show that the property of observational completeness is strictly stronger than informational completeness, thus constituting a new "bureau of standards" in terms of DDI reconstruction. To this aim we show that, for systems with (hyper)spherical state space such as the classical and quantum bits, the only observationally complete simplex is the regular simplex, that is, the *symmetric* informationally complete (SIC) one [35–40]. Data-driven inference and reconstruction, hence, naturally lead to the notion of SIC apparatus *by looking only at the set of outcome distributions*, thus providing a completely new viewpoint on the discussion about SIC apparatus and their "natural occurrence" in quantum theory.

The structure of the paper follows the above discussion. In the first section we introduce data-driven inference as the inference of the minimal range consistent with the observed distributions, and we show that the inferred range is unique for any set of outcome distributions only for systems with (hyper)spherical state space. We also prove that the range of a measurement identifies such a measurement up to gauge symmetries. In the second section we introduce the property of observational completeness and show that it represents a strictly stronger condition than informational completeness. For systems with (hyper)spherical state space, we show that the minimal observationally complete set of states happens to be SIC. In the third section we implement a protocol for the machine learning of qubit measurements based on data-driven inference and reconstruction, and we test it on data generated by the IBM Q Experience quantum computer.

II. DATA-DRIVEN INFERENCE

Let us consider an experimental setup involving two boxes equipped with m buttons and n light bulbs, respectively. This situation is depicted in Fig. 1. At each run of the experiment, a theoretician, say Alice, presses button x and observes outcome y. She records the vectors $\{\mathbf{p}_x \in \mathbb{R}^n\}$, whose yth entry is the frequency of outcome y given input x.

We address the problem of inferring all the measurements $\hat{M}$ that are *self-consistent* and *minimal* for observed frequencies $\{\mathbf{p}_x\}$, in an i.i.d. hypothesis for $\hat{M}$. To formalize this idea, notice that any n-outcome measurement M induces a linear transformation from the state space $\mathbb{S}$ (the set of all states available to the system) to the space of outcome distributions $\mathbb{R}^n$. The range of such a transformation, denoted by $M(\mathbb{S})$, represents the distributions compatible with measurement M.

---

[2]On the contrary, we *need* to assume that the unknown measurement to be reconstructed remains the same during the entire experiment—otherwise the problem of reconstruction would not even be well defined.

FIG. 1. The experimental setup consists of two uncharacterized boxes equipped with m buttons and n light bulbs, respectively. At each run, Alice presses button x and observes outcome y, thus recording vectors $\{\mathbf{p}_x \in \mathbb{R}^n\}$, whose yth entry is the frequency of outcome y given input x.
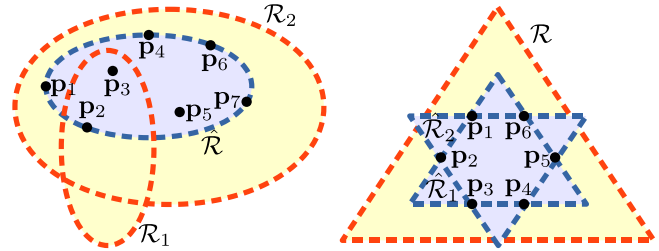

FIG. 2. A schematic representation of the space of distributions when $n = 3$ (three outcomes, and hence the probability simplex is two dimensional). Left: When the state space $\mathbb{S}$ is a sphere, the data-driven inference defined by Eq. (1) returns the minimum volume-enclosing ellipsoid $\hat{\mathcal{R}}$, which always exists unique, for the given set $\{\mathbf{p}_x\}$. Range $\mathcal{R}_1$ is not consistent with $\{\mathbf{p}_x\}$, while $\mathcal{R}_2$, although consistent, is not minimum volume. Right: When the state space $\mathbb{S}$ is a simplex, the data-driven inference defined by Eq. (1) returns the set of minimum volume-enclosing simplices (as these are generally not unique) $\hat{\mathcal{R}}_1$ and $\hat{\mathcal{R}}_2$ for the given set $\{\mathbf{p}_x\}$. Range $\mathcal{R}$ is not minimum volume.

Hence, given some prior information about a state space $\mathbb{S}$, the inferred linear transformations $\hat{M}$ minimize the volume[3] of their ranges $\hat{M}(\mathbb{S})$ under the self-consistency constraints $\hat{M}(\mathbb{S}) \supseteq \{\mathbf{p}_x\}$.[4] This naturally identifies two steps in the inference of $\hat{M}$: (i) the inference of the (possibly not unique) minimum-volume range $\hat{\mathcal{R}}$ such that $\hat{\mathcal{R}} \supseteq \{\mathbf{p}_x\}$, and (ii) the characterization of all the linear transformations $\hat{M}$ with given range $M(\mathbb{S}) = \hat{\mathcal{R}}$.

Let us start addressing the first step, that is, the inference of the (possibly not unique) minimum-volume range $\hat{\mathcal{R}}$ such that $\hat{\mathcal{R}} \supseteq \{\mathbf{p}_x\}$. Since such an inference is solely driven by the data, we call it *data-driven inference* (see also [41,42]):

Definition 1 (Data-driven inference). For any set $\{\mathbf{p}_x\}$, we denote by $\mathtt{ddi}(\{\mathbf{p}_x\}|\mathbb{S})$ the *data-driven inference* map

$$\mathtt{ddi}(\{\mathbf{p}_x\}|\mathbb{S}) = \mathrm{argmin}_{\mathcal{R} \supseteq \{\mathbf{p}_x\}} \mathcal{V}(\mathcal{R}), \qquad (1)$$

where $\mathcal{V}(\mathcal{R})$ denotes the Euclidean volume of $\mathcal{R}$ and the minimization is over subsets $\mathcal{R} \subseteq \mathbb{R}^n$ corresponding to linear transformations of the state space $\mathbb{S}$ that lie on the affine subspace generated by $\{\mathbf{p}_x\}$.

If the prior information does not specify a single state space $\mathbb{S}$, the minimum in Eq. (1) is meant to run also over all such sets.

An intuitive geometrical interpretation for Definition 1 is illustrated by the following two examples.

First, let $\mathbb{S}$ be a (hyper)sphere Σ. This scenario encompasses the cases of classical and quantum bits, where the set $\mathbb{S}$ is a one-dimensional and a three-dimensional sphere, respectively. In this case, the optimization in Eq. (1) is over affine transformations of a sphere, that is, ellipsoids. It follows [43,44] that the range returned by map $\mathtt{ddi}$ is unique for any input set $\{\mathbf{p}_x\}$, and one has that

$$\mathtt{ddi}(\{\mathbf{p}_x\} \mid \Sigma) = \{\mathtt{MVEE}(\{\mathbf{p}_x\})\}, \qquad (2)$$

where $\mathtt{MVEE}$ denotes the minimum volume-enclosing ellipsoid, which is known to be a convex problem, and for which efficient computing algorithms are available [45]. This situation is illustrated in Fig. 2, left-hand side.

Let us consider now, as a second example, the case in which the state space $\mathbb{S}$ is a regular simplex Δ, as it is the case for classical systems. In this case, the optimization in Eq. (1) is over affine transformations of the simplex Δ, which turn out to be simplices themselves. It is not difficult to find configurations of the set $\{\mathbf{p}_x\}$ of observed distributions, such that the range returned by the map $\mathtt{ddi}$ is not unique. This situation is illustrated in Fig. 2, right-hand side.

In light of the previous examples, it is a natural question to ask for which state spaces $\mathbb{S}$ the range returned by the data-driven inference $\mathtt{ddi}$ is always unique. The following Theorem, proved in Appendix B, answers such a question.

Theorem 1. The range returned by the data-driven inference $\mathtt{ddi}(\{\mathbf{p}_x\}|\mathbb{S})$ is unique for any $\{\mathbf{p}_x\}$, if and only if the state space $\mathbb{S}$ is a (hyper)sphere.

This result can be lifted to the level of a principle, singling out spherical state spaces $\mathbb{S}$, such as those of the classical and quantum bits, as those for which the map $\mathtt{ddi}$ always returns a unique range. This principle rules out theories with more exotic elementary systems, such as PR boxes [25] for which inference is not always unique. In this case, we speak of an *epistemic principle*, that is, a constraint on the state space seen as the hypothesis used by the observer as the base of their inference.

Let us now move on to the second step mentioned above, that is, the characterization of all the linear transformations M with range $M(\mathbb{S})$ equal to a given inferred one $\hat{\mathcal{R}}$. Notice first that any transformation U that leaves the state space $\mathbb{S}$ invariant, that is such that $U(\mathbb{S}) = \mathbb{S}$, does not affect the range $M(\mathbb{S})$, that is $M(U(\mathbb{S})) = M(\mathbb{S})$. We refer to any such a transformation as a *gauge symmetry*. The following Theorem shows that accuracy up to gauge symmetries is indeed the optimal accuracy in the characterization of any informationally complete (i.e., invertible) M, given its range.
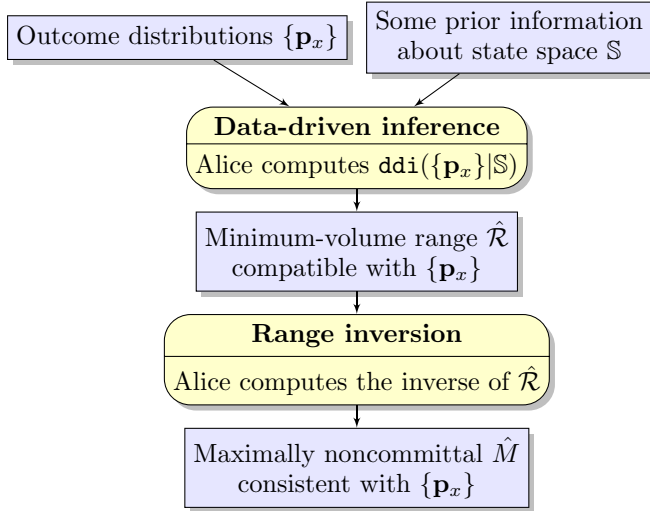
---

[3]Notice that the choice of volume as an objective function is independent of linear transformations, since for any body the volume under any linear transformation changes by a factor that only depends on the linear transformation, and not on the body itself. Nonlinear transformations are not allowed, as they do not preserve the linear structure of the underlying space (state or effect space or probability space).

[4]Notice that not any linear transformation corresponds to a legitimate measurement. In the case in which none of the inferred $\hat{M}$ is a legitimate measurement, the inference fails, in the sense that Alice declares that either the data are insufficient or that the assumption of the state space $\mathbb{S}$ is inconsistent.

FIG. 3. Algorithmic representation of data-driven inference.

FIG. 4. The same experimental setup as in Fig. 1, although this time the state preparator $\mathcal{S}$ is built by Bob, with the aim of enabling Alice to correctly infer measurement range $M(\mathbb{S})$, that is, for the inferred range $\hat{\mathcal{R}}$, one has $\hat{\mathcal{R}} = M(\mathbb{S})$.

(For noninformationally complete M, the statement, although conceptually similar, becomes technically more involved, so we postpone the general statement and its proof to Appendix C.)

Theorem 2. For any given state space $\mathbb{S}$, the range $M(\mathbb{S})$ of any informationally complete M identifies M up to gauge symmetries.

Although Theorem 2 is valid for any state space $\mathbb{S}$, for the sake of concreteness let us revisit our running example where $\mathbb{S}$ is a (hyper)sphere Σ, in particular a qubit.

In the qubit case, the gauge symmetries correspond to unitary and antiunitary transformations in the Hilbert space, hence, the range $M(\mathbb{S})$ identifies linear transformation M up to unitary and antiunitary transformations. However, Theorem 2 only guarantees the existence of such an identification, without providing an explicit construction. Reference [41] fills this gap by explicitly deriving all the linear transformations that correspond to any given (hyper)ellipsoidal range. Such a result is provided with a new simpler proof as Proposition 1 of Appendix C.

An algorithmic representation of data-driven inference is given in Fig. 3.

III. DATA-DRIVEN RECONSTRUCTION

In the previous section we introduced a principle of self-consistent minimality to guide the inference of a measurement given a set of observed outcome distributions. In this section we consider the case in which the boxes with buttons and lights in Fig. 1 describe a physical state-preparator $\mathcal{S}$ and a physical measurement M, respectively. This situation is illustrated in Fig. 4.

The state-preparator $\mathcal{S}$ is built by an experimentalist, say Bob, with the aim of enabling Alice to correctly infer the measurement range $M(\mathbb{S})$ through DDI, that is, the inferred range $\hat{\mathcal{R}}$ satisfying $\hat{\mathcal{R}} = M(\mathbb{S})$, in the limit in which Alice presses each button infinitely many times. This task shares similarities with conventional measurement tomography, with the major difference that, in the latter, full knowledge of the
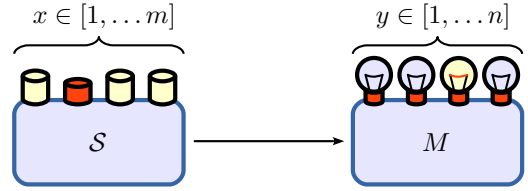
state-preparator $\mathcal{S}$ is pivotal for Alice's data analysis, whereas DDI solely depends on the observed outcome distributions and the knowledge of the state space $\mathbb{S}$.

Since it is sufficient to show that Bob is able to construct one such state preparator, we can assume, without loss of generality, that each button of the state preparator emits always the same state at each press, and that different buttons are associated with different states. Hence, the state-preparator $\mathcal{S}$ can be mathematically described as a set of states.

The probabilities $\{\mathbf{p}_x\}$ Alice observes are the image of the states in $\mathcal{S}$, that is, $\{\mathbf{p}_x\} = M(\mathcal{S})$. Correct inference imposes then that $M(\mathcal{S})$ contains all the statistical information that is available in the measurement range $M(\mathbb{S})$, and that such information can be extracted by data-driven inference. We call *observationally complete* (OC) any state preparator that allows for the correct inference of measurement M.

Definition 2 (Observational completeness). A set of states $\mathcal{S}$ is said to be *observationally complete* for measurement M whenever

$$\mathtt{ddi}(M(\mathcal{S})|\mathbb{S}) = \{M(\mathbb{S})\}. \tag{3}$$

If the set is not OC, the set of inferred measurements may not contain the correct measurement, but measurements with ranges smaller than the correct measurement, or even measurements inequivalent (modulo gauge symmetries) to it.

Notice that observational completeness (OC-ness) plays an analogous role for data-driven reconstruction as informational completeness (IC-ness) plays for conventional measurement tomography. However, IC sets of states allow for the correct tomographic reconstruction of *any* measurement, while OC sets of states apparently depend on the measurement to be inferred.

Is this really the case? It turns out that, as long as the measurement is IC, by bypassing the linear transformation M in Eq. (3) one obtains a condition equivalent to Eq. (3), as stated in the following theorem:

Theorem 3. A set $\mathcal{S}$ of states is OC complete for any IC measurement, if and only if

$$\mathtt{ddi}(\mathcal{S}|\mathbb{S}) = \{\mathbb{S}\}. \tag{4}$$

Notice that in Eq. (4) the map $\mathtt{ddi}$ is applied to a set of states, as opposed to a set of probability distributions as it was the case so far.

Hence, any set $\mathcal{S}$ of states that is OC for some IC measurement, is also OC for *any other* IC measurement. Moreover, Eq. (4) provides a characterization of any such a set $\mathcal{S}$ in closed form, namely, in a form which only depends on $\mathcal{S}$
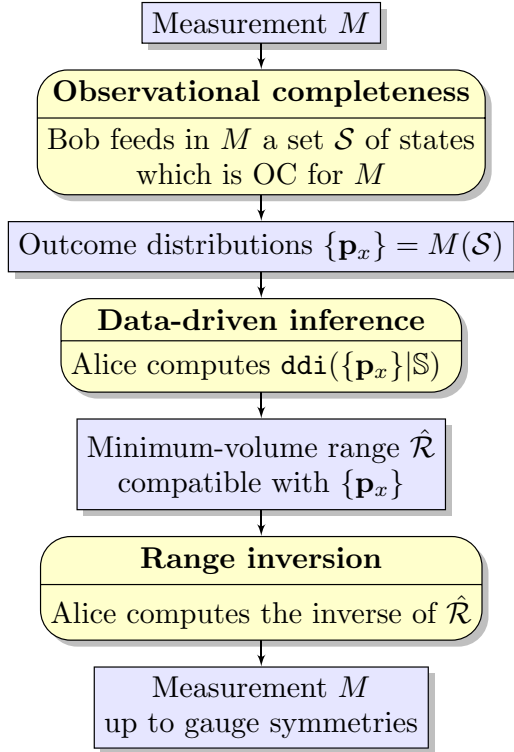
FIG. 5. Algorithmic representation of data-driven reconstruction.

FIG. 6. A two-dimensional representation of the spherical state space $\mathbb{S}$ (orange line) and of the minimum volume-enclosing ellipsoid (blue line) for the simplicial set $\mathcal{S}$ of states (black vectors). Left: When $\mathcal{S}$ coincides with the regular simplex Δ, its minimum-volume enclosing ellipsoid coincides with the sphere Σ, hence, due to Theorem 3, $\mathcal{S}$ is observationally complete. Right: When $\mathcal{S}$ does not coincide with the regular simplex Δ, its minimum volume-enclosing ellipsoid does *not* coincide with the sphere Σ, hence, due to Theorem 3, $\mathcal{S}$ is not observationally complete.

alone, in contrast with Definition 2 that also depends on M. Such a characterization can be used to readily check if any given set $\mathcal{S}$ is OC for any IC measurement.

More generally, even if M is not an IC measurement, one can write a condition equivalent to Eq. (3) [and conceptually analogous to Eq. (4), just technically more involved] that depends on M only through its support. Hence, any set $\mathcal{S}$ of states that is OC for M is also OC for any other measurement with the same support. In other words, it is *universally* OC on such a support. A fully general version of Theorem 3 is provided in Appendix D.

Notice that, while OC sets of states are universal within a given subspace, IC sets of states are universal for any subspace and *all* of its subspaces. In fact, it is easy to see that the only set of states that is OC for *any* measurement M is trivially the state space $\mathbb{S}$.

An algorithmic représentation of data-driven reconstruction is given in Fig. 5.

Although Theorem 3 is valid for any state space $\mathbb{S}$, for the sake of concreteness let us revisit our running example where $\mathbb{S}$ is a (hyper)sphere.

In light of Eq. (2), Theorem 3 states that a set $\mathcal{S}$ of states is OC for IC measurements if and only if its minimum volume-enclosing ellipsoid coincides with the (hyper)sphere Σ, that is

$$\mathtt{MVEE}(\mathcal{S}) = \Sigma. \qquad (5)$$

In Appendix D (Proposition 2), as a consequence of Ref. [46], we show that such a condition is satisfied when $\mathcal{S}$ is a regular simplex. This situation is illustrated in Fig. 6, left-hand side. Moreover, we show that, as a consequence of Ref. [47], also
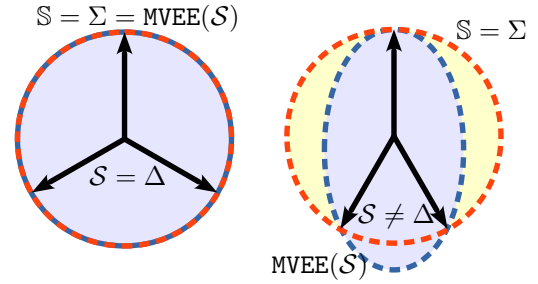
the converse is true, namely, that such a condition is violated whenever $\mathcal{S}$ is an irregular simplex (see Fig. 6, right-hand side). Hence, for simplices, OC-ness is equivalent to symmetric IC-ness and, therefore, the SIC set of states is the minimal (in terms of cardinality) OC set. This provides an operational interpretation of symmetric informational completeness in terms of data-driven inference and reconstruction. It is tempting to conjecture that this equivalence holds for any quantum system, not just the qubit.

Although it is easy to see that any set of states which is OC on a subspace, is also IC on that subspace, the previous example shows that the vice versa is not true. Indeed, any simplex, whether regular or not, is trivially IC on its support. Hence, OC-ness is a strictly stronger condition than IC-ness. In this sense, OC-ness defines a new "bureau of standards" in terms of data-driven inference and reconstruction.

IV. MACHINE LEARNING OF QUBIT MEASUREMENTS

In this section we first provide an algorithm for the data-driven inference of measurements of any system with (hyper)spherical state space, and then we apply it to the data-driven reconstruction of a qubit measurement implemented with the IBM quantum computer.

Let us begin by discussing the algorithm for the data-driven inference. As shown by Eq. (2), for (hyper)spherical state space $\mathbb{S} = \Sigma$ the data-driven inference algorithm $\mathtt{ddi}$ can be written in terms of a minimum volume enclosing ellipsoid algorithm $\mathtt{MVEE}$. Efficient algorithms for the computation of $\mathtt{MVEE}$ (quadratic in the dimension and linear in the number of points) can readily be found [45,48], but in general assume that the affine space generated by the input set coincides with the full linear space. Hence, given a set $\{\mathbf{p}_x\}$, to compute $\mathtt{ddi}(\{\mathbf{p}_x\}|\Sigma)$, one proceeds as follows:

(1) Find an isometry V such that $V^T V = \mathbb{1}$ and $V V^T$ is the projector on the linear subspace homogeneous to the affine subspace generated by $\{\mathbf{p}_x\}$.

(2) Compute $\mathtt{MVEE}(V^T \{\mathbf{p}_x - \mathbf{p}_0\})$, where by construction the affine subspace generated by $V^T \{\mathbf{p}_x - \mathbf{p}_0\}$ is now full
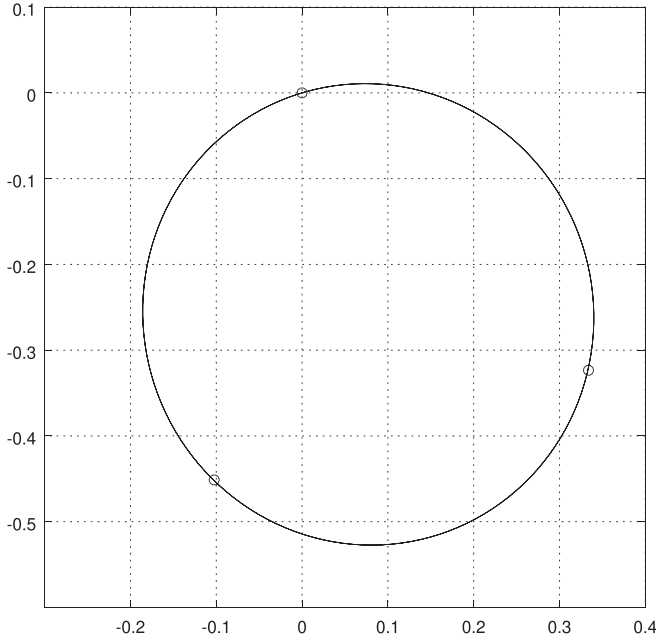
FIG. 7. The three points $V^T(\mathbf{p}_x - \mathbf{p}_0)$ (circles) for $x = 0, 1, 2$, where $\mathbf{p}_x$'s are given by Eq. (6), and their minimum-volume enclosing ellipsoid, according to the second step of our algorithm for the data-driven inference of qubit measurements.

dimensional, thus obtaining an ellipsoid in the form $(\tilde{\mathbf{p}} - \tilde{\mathbf{t}})^T \tilde{Q}^{-1} (\tilde{\mathbf{p}} - \tilde{\mathbf{t}}) \leqslant 1$.

(3) The output of $\mathtt{ddi}(\{\mathbf{p}_x\}|\Sigma)$ is the singleton given by the ellipsoid

$$(\mathbb{1} - QQ^+)(\mathbf{p} - \mathbf{t}) = 0,$$
$$(\mathbf{p} - \mathbf{t})^T Q^+ (\mathbf{p} - \mathbf{t}) \leqslant 1,$$

where $Q := V\tilde{Q}V^T$ and $\mathbf{t} := V\tilde{\mathbf{t}} + \mathbf{p}_0$.

We tested our algorithm on the set $\{\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2\}$ as follows (later on we show how this data were obtained):

$$\mathbf{p}_0 \simeq \begin{bmatrix} 0.48 \\ 0.05 \\ 0.24 \\ 0.23 \end{bmatrix}, \quad \mathbf{p}_1 \simeq \begin{bmatrix} 0.21 \\ 0.32 \\ 0.42 \\ 0.05 \end{bmatrix}, \quad \mathbf{p}_1 \simeq \begin{bmatrix} 0.17 \\ 0.33 \\ 0.11 \\ 0.39 \end{bmatrix}. \quad (6)$$

The minimum-volume enclosing ellipsoid for the set $\{V^T(\mathbf{p}_x - \mathbf{p}_0)\}$, as given in the second step of our algorithm, is depicted in Fig. 7. The correlation matrix Q and the vector t, as reconstructed by the third step of our algorithm, are given by

$$Q \simeq 10^{-2} \begin{bmatrix} 3.8 & -3.6 & 0 & -0.1 \\ -3.6 & 3.4 & 0.2 & 0 \\ 0 & 0.2 & 3.3 & -3.4 \\ -0.1 & 0 & -3.4 & 3.6 \end{bmatrix}, \quad \mathbf{t} \simeq \begin{bmatrix} 0.29 \\ 0.23 \\ 0.26 \\ 0.22 \end{bmatrix}. \tag{7}$$

As a consequence of Theorem 2, through range inversion this allows for the data-driven inference of the POVM up to gauge symmetries, that is, an overall unitary or antiunitary transformation. In particular, as shown in Ref. [41] (an alternative, more compact proof is given in given in Proposition 1 of Appendix C), any measurement $\{\pi_y\}$ compatible with
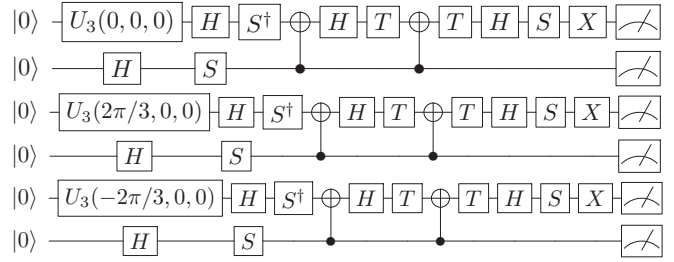

FIG. 8. The orange gate (top left) represents the preparation of the trine ensemble (from top to bottom, states ϕ_0, ϕ_1, and ϕ_2). All other gates represent a Naimark extension of a measurement consisting of two mutually unbiased basis.

correlation matrix Q and vector t can be obtained by inverting the following system:

$$Q_{x,y} = \tfrac{1}{2}\mathrm{Tr}[\pi_x \pi_y] - \tfrac{1}{4}\mathrm{Tr}[\pi_x]\mathrm{Tr}[\pi_y], \quad t_x = \tfrac{1}{2}\mathrm{Tr}[\pi_x].$$

Let us now discuss the data-driven reconstruction, that is, the way in which the data in Eq. (6) was generated. Such data were generated by the IBM Q Experience quantum computer. The ideal circuit consists of the preparation of a trine set $\{\phi_x\}$ of states, which, as observed in Fig. 6, is observationally complete for any real measurement, and the measurement of two mutually unbiased basis, as depicted in Fig. 8. This allows one to compare the correlation matrix Q and the vector $\mathbf{t}$ given in Eq. (7) (from which one can recover the inferred measurement up to gauge symmetries through range inversion) with the ideal ones, given by

$$Q = \frac{1}{16} \begin{pmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \quad \mathbf{t} = \frac{1}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

An analysis of the noise affecting the IBM back end is however out of the scope of this paper.

V. CONCLUSION

In this work we introduced data-driven inference as a rule to output the maximally noncommittal measurement consistent with a set of observed distributions. We showed that the inference is possible in principle up to gauge symmetries, that is, symmetries of the state space of the system at hand, and that this accuracy limit is achieved for (hyper)spherical state spaces. Then we considered the task of reconstructing an unknown measurement via DDI. To this aim we introduced observationally complete sets of states, as those enabling a correct inference *universally*, that is, for *any* unknown measurement on a given support. Deriving a closed-form characterization of observational completeness allowed us to show that, while observational completeness is a strictly stronger condition than informational completeness, in the case of (hyper)spherical state space OC-ness with minimum number of states is equivalent to SIC-ness. In this way, the protocol of data-driven inference provides SIC sets with a novel, entirely operational interpretation.

APPENDIX A: GENERAL FRAMEWORK

While, for the sake of clarity, in the previous sections we focused on the data-driven inference of measurements, in these Appendixes we extend the formalism to encompass the case of data-driven inference of families of states, thus justifying the word "devices" in our title. Finally, rather than restricting the presentation to the quantum case, here we consider general probabilistic theories.

A physical system can be defined by giving a set of states and a set of effects, representing, respectively, the preparations and the observations of the system. An effect $a \in \mathbb{E}$ is a linear map that takes a state ρ as an input and outcomes a probability $p(a|\rho) := a(\rho)$. Since randomization of different experimental setups is in itself another valid experiment, it is natural to endow the set of states and the set of effects with a linear structure and allow for any convex combination of states and effects. By linear extension, it is also natural to introduce the real vector spaces generated by any real linear combinations of states and effects. Restricting to the finite dimensional case, the linear space of states and the linear space of effects are dual to each other and both isomorphic to $\mathbb{R}^\ell$ for some natural number ℓ which is called the *linear dimension* of the system.

We assume that the physical theory is *causal*. A probabilistic theory is causal if there exists a unique *deterministic effect* $e \in \mathbb{E}$, and *deterministic states* are those states such that $e(\rho) = 1$. Therefore, states can always be normalized as $\overline{\rho} := \rho/e(\rho)$ and every state is proportional to a deterministic one. For this reason, the full set of states of any causal theory is completely specified by the set of deterministic states. We denote by $\mathbb{S}$ the set of normalized (or deterministic) states of the theory and by $\mathbb{E}$ the set of effects of the theory.

By choosing an arbitrary basis, we can give a geometric representation of $\mathbb{S}$ and $\mathbb{E}$ as a subset of $\mathbb{R}^\ell$: $\mathbb{S}$ will be a convex set contained in a strictly affine $(\ell - 1)$-dimensional subspace, while $\mathbb{E}$ will be a "bicon-ish" shaped solid (see Fig. 9).

Measurements are a family of effects $\{a_y\}_{y=1}^n$ such that $\sum_{y=1}^n a_y = e$, $\forall \rho \in \mathbb{S}$. Since any state can be regarded as a vector in $\mathbb{R}^\ell$, any measurement induces a linear map $M \in \mathfrak{M}_{n \times \ell}$ defined as follows:

$$M : \mathbb{R}^l \to \mathbb{R}^n, \quad \rho_x \mapsto \mathbf{p}_x, \quad p_x^y := a_y(\rho_x)$$

(each row of $M \in \mathfrak{M}_{n \times \ell}$ corresponds to an effect). Any state is mapped into a point in $\mathbb{R}^n$ (see the top of Fig. 10). Analogously, since any effect a_x corresponds to a vector in $\mathbb{R}^\ell$,
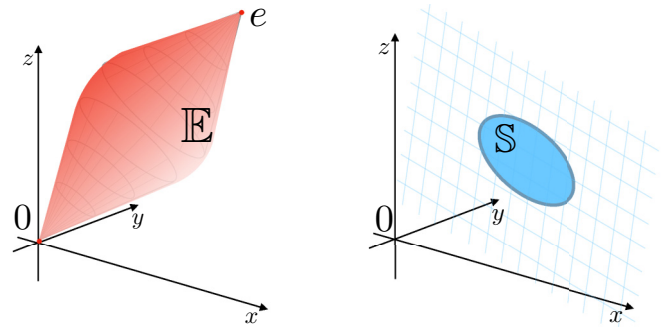

FIG. 9. Vectorial representation of the system $(\mathbb{S}, \mathbb{E})$ of linear dimension $\ell = 3$. Both the set of normalized states $\mathbb{S}$ and the set of effects $\mathbb{E}$ can by represented as convex subsets of the vector space $\mathbb{R}^3$. Left: The convex set of effects $\mathbb{E}$ corresponds to a truncated cone. The truncated cone contains the two points $(0,0,0)$ and $(1,1,1)$, which correspond to the null effect and to the deterministic effect e, respectively. Right: The convex set of normalized states $\mathbb{S}$. This set contains all states ρ such that $e(\rho) = 1$, and is then merged in an affine space with dimension $\ell - 1$ (in this case a two-dimensional space).

any family of states $\{\rho_y\}_{y=1}^n$ induces a linear map defined as follows:

$$R : \mathbb{R}^l \to \mathbb{R}^n, \quad a_x \mapsto \mathbf{p}_x, \quad p_x^y := \rho_y(a_x)$$

(each row of $R \in \mathfrak{M}_{n \times \ell}$ corresponds to a state). Any effect is mapped in a point in $\mathbb{R}^n$ (see the bottom of Fig. 10).

For example, in quantum theory, any system is associated with a d-dimensional Hilbert space $\mathcal{H}$, states and effects are represented by positive semidefinite operators on $\mathcal{H}$, and conditional probabilities are given by the Born rule: $a_y(\rho_x) := \mathrm{Tr}[a_y \rho_x]$. States and effects are represented by vectors in
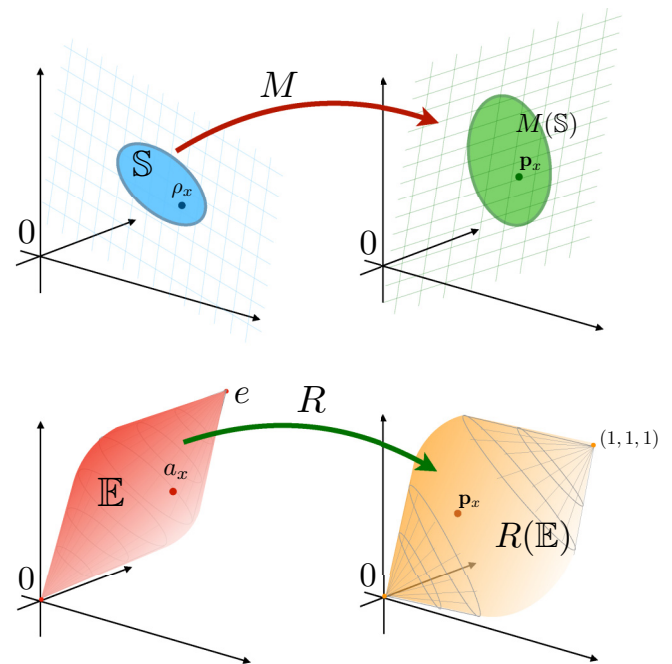

FIG. 10. Top: The linear map M corresponding to a measurement. Bottom: The linear map R corresponding to a family of states.

the real space $\mathbb{R}^{\ell}$ with $\ell = d^2$. As an example, any qubit ($d = 2$) normalized state $\rho = \frac{1}{2}\sum_{i=0}^{3} r^i \sigma_i$ is univocally associated with a vector $(r^0, r^1, r^2, r^3)^T \in \mathbb{R}^4$. Here the condition $r^0 = 1$ guarantees that $\text{Tr}[\rho] = 1$, while the condition $|r| \leqslant 1$ for Bloch vector $\mathbf{r} := (r^1, r^2, r^3)^T$, where $\{\sigma_i\}_{i=1}^{3}$ are the Pauli matrices, guarantees that $\rho \geqslant 0$. The set of normalized states, identified by the constraint $r^0 = 1$, is geometrically represented by a sphere (the Bloch sphere) contained in a three-dimensional strictly affine subspace of $\mathbb{R}^4$.

As a second example, in classical theory the set of normalized states of any system with linear dimension ℓ is an ℓ simplex, e.g., a segment for the *bit* system $\ell = 1$ and a triangle for the *trit* system $\ell = 2$.

In the literature, toy models have been proposed whose convex set of states (and effects) differs from both the quantum and the classical ones. The most notable example are the PR boxes, whose convex set of states of linear dimension $\ell = 3$ is a square.

APPENDIX B: DATA-DRIVEN INFERENCE

The *data-driven inference* (DDI) of quantum measurements, presented in the main text, is a protocol that allows us to infer a preferred (according to a maximally noncommittal criterion) measurement from a set of data interpreted as the outcome distributions of an experiment. The maximally noncommittal criterion is very natural: among the set of measurements whose range includes the experimental points, we choose those with minimum-volume range in the space of outcome distributions. Hence, the algorithmic idea at the basis of the data-driven inference of quantum measurements is very simple: (i) the first step is the search of the minimum-volume enclosing ranges for a given set of points, and (ii) the second step is the search of the measurements that are able to reproduce such ranges as the outcome distributions of an experiment.

It is intuitive that in an analogous way one can define data-driven inference of quantum states. Moreover, due to its purely geometrical nature, the idea of data-driven inference of physical devices is not anchored to the quantum formalism, but can be defined in the same way in any possible probabilistic theory. Within this perspective, here we define the data-driven inference of measurements and of states in the framework of general probabilistic theories [24–31].

1. Inference of measurements

A setup comprising two boxes, one equipped with n bulbs and the other equipped with m buttons, is given. At each run of the experiment a theoretician, say Alice, presses button x and records which bulb y lights up. She iterates this procedure many times, recording the frequencies $\{\mathbf{p}_x\}$ whose yth element is the probability of outcome y given input x. This situation is illustrated in the upper part of Fig. 11.

The aim of data-driven inference is to infer the maximal noncommittal linear maps $M : \mathbb{R}^l \rightarrow \mathbb{R}^n$ consistent with $\{\mathbf{p}_x\}$, that is, the linear map M that minimize the volume of $M(\mathbb{S})$ such that the range $M(\mathbb{S})$ contains the distributions $\{\mathbf{p}_x\} \subseteq \mathbb{R}^n$. We recall that $\mathbb{S}$ is the set of all states of the system of linear dimension ℓ. As already noticed in the main text, not
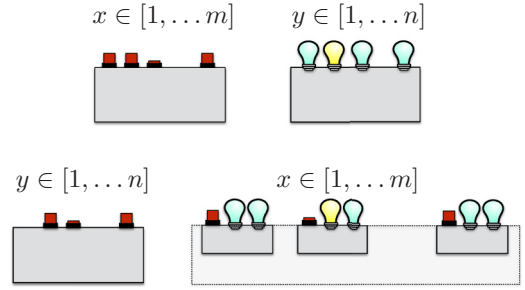

FIG. 11. Top: Setup for the inference of measurements. Bottom: Setup for the inference of families of states.

any linear transformation corresponds to a legitimate measurement: in case of a nonphysical inference M, failure is declared and a larger set $\{\mathbf{p}_x\}$ is required. This definition of the problem naturally identifies two steps: (i) inferring the (possibly nonunique) minimum-volume range $\hat{\mathcal{R}}$ consistent with $\{\mathbf{p}_x\}$, and (ii) finding the measurements $\hat{M}$ whose range is $\hat{\mathcal{R}}$.

In the following definition we formalize the first of these two steps, that is, the inference process that consists of finding the minimum-volume range $M(\mathbb{S})$ such that $M(\mathbb{S}) \supseteq \{\mathbf{p}_x\}$.

Definition 1 (Data-driven inference of measurements). For any set $\{\mathbf{p}_x\} \subseteq \mathbb{R}^n$, we denote with $\text{ddi}(\{\mathbf{p}_x\} \,|\, \mathbb{S})$ the *data-driven inference* map

$$\text{ddi}(\{\mathbf{p}_x\} \,|\, \mathbb{S}) = \underset{\substack{M(\mathbb{S}) \\ \{\mathbf{p}_x\} \subseteq M(\mathbb{S}) \subseteq \text{aff}\{\mathbf{p}_x\}}}{\text{argmin}} \mathcal{V}(M(\mathbb{S})), \quad \text{(B1)}$$

where $\mathcal{V}(\mathcal{R})$ denotes the Euclidean volume of $\mathcal{R}$ and the minimization is over subsets $M(\mathbb{S})$ corresponding to linear transformations of the set of states $\mathbb{S}$ that lie on the affine subspace generated by $\{\mathbf{p}_x\}$.

In general the outcome of the DDI map is not unique, and the map returns a set of ranges. We notice that, if the available prior information does not identify a unique set $\mathbb{S}$ of states, the set $\mathbb{S}$ itself can be considered as part of the optimization problem in the above definition, by taking the minimum of Eq. (B1) over any possible $\mathbb{S}$.

We can now derive the main result of this section that establishes the special role played by (hyper)ellipsoidal sets $\mathbb{S}$ of states in the context of the DDI of measurements.

Before stating the main theorem we need the following definition.

Definition 2 (U-symmetric set). Given a set of transformations $\mathcal{U}$, and a set $\mathcal{X} \subseteq \mathbb{R}^k$, we say that $\mathcal{X}$ is $\mathcal{U}$ symmetric if $U(\mathcal{X}) := \{U\mathbf{x} | \mathbf{x} \in \mathcal{X}\} = \mathcal{X}$ for any $U \in \mathcal{U}$, where we have chosen a k-dimensional representation of the set $\mathcal{U}$.

Theorem 1. Given a set of states $\mathbb{S}$ the following conditions are equivalent:

(1) $\text{ddi}(\mathcal{X}|\mathbb{S})$ is a singleton for any $\mathcal{X} \subseteq \mathbb{R}^n$ and any $n \in \mathbb{N}$.

(2) $\text{ddi}(\mathcal{X}|\mathbb{S})$ is $\mathcal{U}$ symmetric for any $\mathcal{U}$-symmetric $\mathcal{X}$, where $\mathcal{U}$ is a set of orthogonal matrices.

(3) $\mathbb{S}$ is a (hyper)ellipsoid.

Proof. Let us prove each implication separately:

$1 \Rightarrow 2$

Consider a $\mathcal{U}$-symmetric $\mathcal{X}$, and suppose by absurd that $\mathtt{ddi}(\mathcal{X}|\mathbb{S})$ is a singleton $\mathcal{Y}$, but $\mathcal{Y}$ is not $\mathcal{U}$ symmetric. Then there exists $U \in \mathcal{U}$ such that $\mathcal{Y}' := \{U\mathbf{y} \mid \mathbf{y} \in \mathcal{Y}\} \neq \mathcal{Y}$. Since $\mathcal{Y}' \supseteq \mathcal{X}$ and $\mathcal{V}(\mathcal{Y}') = \mathcal{V}(\mathcal{Y})$, one has the absurd $\mathcal{Y}' \in \mathtt{ddi}(\mathcal{X}|\mathbb{S})$.

$2 \Rightarrow 3$

The implication follows from Lemma 1, by taking $\mathcal{X}$ to be a sphere and observing that $\{M(\mathbb{S}) \mid M \in \mathfrak{M}_{n \times \ell}\}$ contains all spheres if and only if $\mathbb{S}$ is a (hyper)ellipsoid.

$3 \Rightarrow 1$

Follows immediately from Lemma 1, due to John's uniqueness theorem for minimum-volume enclosing ellipsoids [43].

Notice that for (hyper)ellipsoidal $\mathbb{S}$, for any $n \in \mathbb{N}$ and any $\mathcal{X} \subseteq \mathbb{R}^n$, one has

$$\mathtt{ddi}(\mathcal{X}|\mathbb{S}) = \{\mathtt{MVEE}(\mathcal{X})\},$$

where $\mathtt{MVEE}(\mathcal{X})$ denotes the minimum-volume enclosing ellipsoid [45] for $\mathcal{X}$.

2. Inference of states

A setup comprising several boxes is given. One box is equipped with n buttons, and the remaining m boxes are equipped with one button and two light bulbs each. At each run of the experiment, Alice presses button y of the former box, and selects box x among the remaining boxes by pressing its button. She iterates this procedure many times, recording the frequencies $\{\mathbf{p}_x\}$ whose yth element is the probability of the first bulb of box x to light up given y. This situation is illustrated in the bottom part of Fig. 11.

The aim of data-driven inference is to infer the maximal noncommittal linear maps $R : \mathbb{R}^l \to \mathbb{R}^n$ consistent with $\{\mathbf{p}_x\}$, that is, the linear maps R that minimize the volume of $R(\mathbb{E})$ such that the range $R(\mathbb{E})$ contains the distributions $\{\mathbf{p}_x\} \subseteq \mathbb{R}^n$. We recall that $\mathbb{E}$ is the set of all effects of the system of linear dimension ℓ. Notice that not any linear map R corresponds to a set of physical states: in case of a nonphysical inference R, failure is declared and a larger set $\{\mathbf{p}_x\}$ is required. This definition of the problem naturally identifies two steps: (i) inferring the (possibly nonunique) minimum-volume range $\hat{\mathcal{R}}$ consistent with $\{\mathbf{p}_x\}$, and (ii) finding the linear transformations $\hat{R}$ whose range is $\hat{\mathcal{R}}$.

In the following definition we formalize the first of these two steps, that is, the inference process that consists of finding the minimum-volume range $R(\mathbb{E})$ such that $R(\mathbb{E}) \supseteq \{\mathbf{p}_x\}$.

Definition 3 (Data-driven inference of states). For any set $\{\mathbf{p}_x\} \subseteq \mathbb{R}^n$, we denote with $\mathtt{ddi}(\{\mathbf{p}_x\} \mid \mathbb{E})$ the *data-driven inference* map

$$\mathtt{ddi}(\{\mathbf{p}_x\} \mid \mathbb{E}) = \underset{\substack{R(\mathbb{E}) \\ \{\mathbf{p}_x, \mathbf{0}, \mathbf{1}\} \subseteq R(\mathbb{E}) \subseteq \mathrm{span}\{\mathbf{p}_x, \mathbf{1}\} \\ R(e) = \mathbf{1}}}{\mathrm{argmin}} \mathcal{V}(R(\mathbb{E})), \quad (\mathrm{B2})$$

where the minimization is over subsets $R(\mathbb{E}) \subseteq \mathbb{R}^n$ corresponding to linear transformations of the set of effects $\mathbb{E}$ that lie on the real span generated by $\{\mathbf{p}_x\} \cup \{\mathbf{1}\}$, and the two points $R(0) = \mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{R}^n$, $R(e) = \mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$ are the images of the null effect $0 \in \mathbb{E}$ and of the deterministic $e \in \mathbb{E}$ effect, respectively [this last condition poses a not trivial linear constraint $R(e) = \mathbf{1}$].

In general the output of the DDI map is not unique, and the map returns a set. We notice that, if the available prior information does not identify a unique set $\mathbb{E}$ of effects, the set $\mathbb{E}$ itself can be considered as part of the optimization problem in the above definition, by taking the minimum of Eq. (B2) over any possible $\mathbb{E}$.

In the definition of DDI of states, the set $\{\mathbf{p}_x\}$ has been extended to include the two points $\mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{R}^n$ and $\mathbf{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$, corresponding to the null $0 \in \mathbb{E}$ and to the deterministic effect $e \in \mathbb{E}$, respectively. This is so because the points $\mathbf{0}$ and $\mathbf{1}$, which could be uncollected by Alice, strongly characterize the geometry of the range $R(\mathbb{E})$ of the linear map R (see also Fig. 10).

There are two main differences between the DDI of states and that of measurements given in Definition 1, when regarded as optimization problems. The first difference is in the set of points where the linear function to be inferred is applied. Indeed, in DDI of states the set $\mathbb{E}$ is not a strictly affine subspace of $\mathbb{R}^l$, while in DDI of measurements the convex set $\mathbb{S}$ is a strictly affine subspace of $\mathbb{R}^l$ of dimension $\ell - 1$ (see also Fig. 10). Moreover, the fixed points of the linear map R in the DDI of states introduce a further linear constraint $R(e) = \mathbf{p}_e$ to the optimization problem. Due to these differences we cannot provide a simple characterization of the DDI of states as for example the one in Theorem 1 for DDI of measurements. A more accurate geometrical analysis of the DDI map for states will be the subject of future research.

APPENDIX C: RANGE INVERSION

In both cases of inference presented above the device to be inferred (either a measurement or a family of states) induces a linear map

$$D : \mathbb{R}^\ell \to \mathbb{R}^n, \quad (\mathrm{C1})$$

for some $n \in \mathbb{N}$.

We denote by $\mathcal{A} \in \mathbb{R}^\ell$ a subset of the domain of the map D (this corresponds to the set of states $\mathcal{S}$ or the set of effects $\mathcal{E}$ in the inference protocol) and we denote by $D(\mathcal{A}) \in \mathbb{R}^n$ the image of $\mathcal{A}$ via the map D (this corresponds to the set of points $\{\mathbf{p}_x\}_{x=1}^m$ collected via the inference protocol):

$$D(\mathcal{A}) := \{D\mathbf{a} \mid \mathbf{a} \in \mathcal{A}\}.$$

Finally we denote the set of all linear transformations of $\mathcal{A}$ into $\mathbb{R}^n$ as

$$\mathcal{D}_n(\mathcal{A}) := \{D(\mathcal{A}) \mid \forall D \in \mathfrak{M}_{n \times \ell}\}. \quad (\mathrm{C2})$$

We can now introduce a notion of equivalence for maps D based on the coincidence of their range $D(\mathbb{A})$.

Definition 4 (Equivalence). For any $n \in \mathbb{N}$ and any $D \in \mathfrak{M}_{n \times \ell}$, we denote with $[D]$ the equivalence class

$$[D] := \{D' \in \mathfrak{M}_{n \times \ell} \mid D'(\mathbb{A}) = D(\mathbb{A})\}.$$

Notice that any two elements of $[D]$ do not necessarily share the same support.

In the main text we referred to any transformation U that leaves a set $\mathbb{A}$ invariant, that is such that $U(\mathbb{A}) = \mathbb{A}$, as a *gauge symmetry*. The following theorem shows that the equivalence class in Definition 4 is fully specified by the symmetries of $\mathbb{A}$.

Theorem 2. For any $n \in \mathbb{N}$ and any $D \in \mathfrak{M}_{n \times \ell}$ one has

$$[D] = \{DU \mid \forall U \text{ s.t. } D^+DU(\mathbb{A}) = D^+D(\mathbb{A})\}. \tag{C3}$$

Proof. The statement can be rephrased as follows. For any $D' \in \mathfrak{M}_{n \times \ell}$ the following conditions are equivalent:

(1) There exists $U \in \mathfrak{M}_{n \times \ell}$ such that $D' = DU$ and $D^+DU(\mathbb{A}) = D^+D(\mathbb{A})$.

(2) $D'(\mathbb{A}) = D(\mathbb{A})$.

Let us prove each implication separately:

$1 \Rightarrow 2$

By hypothesis U is such that $D^+DU(\mathbb{A}) = D^+D(\mathbb{A})$. By multiplying both sides by D from the left one has $DD^+DU(\mathbb{A}) = DD^+D(\mathbb{A})$. Since $D = DD^+D$ one has $DU(\mathbb{A}) = D(\mathbb{A})$. Hence the thesis.

$2 \Rightarrow 1$

Since by hypothesis $D'(\mathbb{A}) = D(\mathbb{A})$, one has $\operatorname{span} D'(\mathbb{A}) = \operatorname{span} D(\mathbb{A})$. Since by hypothesis $\operatorname{span} \mathbb{A} = \mathbb{R}^\ell$, one has $\operatorname{span} D(\mathbb{A}) = \operatorname{rng} D$ and $\operatorname{span} D'(\mathbb{A}) = \operatorname{rng} D'$. Hence, $\operatorname{rng} D = \operatorname{rng} D'$, and thus $DD^+ = D'D'^+$. Then $D' = D'D'^+D' = DD^+D'$. By setting $U := D^+D'$ one has $DU(\mathbb{A}) = D(\mathbb{A})$. By multiplying both sides by D^+ from the left one has $D^+DU(\mathbb{A}) = D^+D(\mathbb{A})$. Hence the thesis. ∎

An immediate corollary of the theorem is:

Corollary 1. For any D such that $D^+D = \mathbb{1}$, due to Eq. (C3) one has that any $D' \in [D]$ is equivalent to D up to a symmetry U of the set $\mathbb{A}$.

Notice that if the map D is for example the linear map associated with a measurement, the corollary states that for any informationally complete measurement M the range $M(\mathbb{S})$ identifies M up to gauge symmetries. This is the statement of Theorem 2 in the main text of the paper.

In case of $(\ell - 1)$-dimensional spherical $\mathbb{S}$ one can also explicitly derive all the linear transformations that correspond to any given (hyper)ellipsoidal range. This is the content of the following proposition.

Proposition 1. For $(\ell - 1)$-dimensional unit-spherical $\mathbb{S}$, for any $n \in \mathbb{N}$ and any $D \in \mathfrak{M}_{n \times \ell}$, let $\mathbf{t} := D\mathbf{u}$, let $T = D(\mathbb{1} - \mathbf{u}\mathbf{u}^T)$, and let $Q = TT^T$. One has

$$D(\mathbb{S}) = \left\{ \mathbf{p} \;\middle|\; \left\{ \begin{array}{l} (\mathbb{1} - QQ^+)(\mathbf{p} - \mathbf{t}) = 0 \\ (\mathbf{p} - \mathbf{t})^T Q^+ (\mathbf{p} - \mathbf{t}) \leqslant 1 \end{array} \right. \right\}.$$

Proof. Due to Lemma 1, without loss of generality we take $\mathbb{S}$ to be the unit sphere centered in $\mathbf{u}$ (centering in $\mathbf{u}$ might apparently require a translation of the sphere on the affine subspace, but a linear transformation suffices due to the Lemma; the inverse linear transformation can be performed on the effects, so without restriction one can consider the states centered in $\mathbf{u}$).

One has $\mathbf{a} \in \mathbb{S}$ if and only if $|\mathbf{a} - \mathbf{u}|_2 \leqslant 1$ and $\mathbf{u}^T\mathbf{a} = 1$. For any $\mathbf{a} \in \mathbb{S}$ one has $D(\mathbf{a}) = \mathbf{t} + T\mathbf{a}$. Hence

$$D(\mathbb{S}) = \{\mathbf{p} = \mathbf{t} + T\mathbf{a} \mid |\mathbf{a} - \mathbf{u}|_2 = 1, \; \mathbf{u}^T\mathbf{a} = 1\}.$$

Solutions of $T\mathbf{a} = \mathbf{p} - \mathbf{t}$ in variable $\mathbf{a}$ exist if and only if

$$TT^+(\mathbf{p} - \mathbf{t}) = \mathbf{p} - \mathbf{t}. \tag{C4}$$

Since T^+T and $\mathbf{u}\mathbf{u}^T$ are orthogonal projectors by construction, solutions are given by

$$\mathbf{a} = T^+(\mathbf{p} - \mathbf{t}) + \lambda\mathbf{u} + (\mathbb{1} - T^+T - \mathbf{u}\mathbf{u}^T)\mathbf{v},$$

for any scalar λ and any vector $\mathbf{v}$.

Condition $\mathbf{u}^T\mathbf{a} = 1$ imposes $\lambda = 1$. For any vector $\mathbf{v}$ such that $|\mathbf{a}|_2 \leqslant 1$, the same condition is also verified for $\mathbf{v} = 0$. Since $D(\mathbf{a})$ is independent of $\mathbf{v}$, without loss of generality we take $\mathbf{v} = 0$.

Therefore one has

$$D(\mathbb{S}) = \left\{ \mathbf{p} \;\middle|\; \left\{ \begin{array}{l} (\mathbb{1} - TT^+)(\mathbf{p} - \mathbf{t}) = 0, \\ (\mathbf{p} - \mathbf{t})^T T^{+T} T^+ (\mathbf{p} - \mathbf{t}) \leqslant 1. \end{array} \right. \right\}.$$

By the elementary properties of the Moore-Penrose pseudoinverse, one immediately has that $TT^+ = QQ^+$ and $T^{+T}T^+ = Q^+$. Hence the statement follows. ∎

APPENDIX D: DATA-DRIVEN RECONSTRUCTION

1. Data-driven reconstruction of measurements

In the protocol of data-driven reconstruction of measurements, an experimentalist, say Bob, is in charge of building the state-preparator $\mathcal{S}$ corresponding to the box equipped with buttons in the upper part of Fig. 12. His aim is to enable Alice to correctly infer measurement M, corresponding to the box with light bulbs, up to the equivalence of Theorem 2. In this case, we say that $\mathcal{S}$ is observationally complete for M.

Definition 5 (Observational complete set of states). Let $\{\mathbb{S}, \mathbb{E}\}$ be a physical system of linear dimension ℓ and $M \in \mathfrak{M}_{n \times \ell}$ be a measurement. A set of states $\mathcal{S} \subseteq \mathbb{S}$ is *observationally complete for M* if and only if

$$\mathtt{ddi}(M(\mathcal{S})|\mathbb{S}) = M(\mathbb{S}), \tag{D1}$$

where $\mathtt{ddi}$ is the data driven inference of measurements.

The following result shows that the notion of observational completeness depends only on the support of M.

Theorem 3. Let $\{\mathbb{S}, \mathbb{E}\}$ be a physical system of linear dimension ℓ and let $\mathcal{S} \subseteq \mathbb{S}$ be a set of states. Let $\mathcal{V}$ be a linear subspace of $\mathbb{R}^\ell$ and let Π denote the projector on $\mathcal{V}$. Then $\mathcal{S}$ is observationally complete for Π if and only if it is observationally complete for any $M \in \mathfrak{M}_{n \times \ell}$ such that $\operatorname{supp} M = \mathcal{V}$, i.e.,

$$\begin{aligned} &\mathtt{ddi}(\Pi(\mathcal{S})|\mathbb{S}) = \Pi(\mathbb{S}) \iff \\ &\mathtt{ddi}(M(\mathcal{S})|\mathbb{S}) = M(\mathbb{S}) \quad \forall M \text{ s.t. } \operatorname{supp} M = \mathcal{V}. \end{aligned} \tag{D2}$$

Proof. We only need to prove the $\Rightarrow$ direction since the opposite one is trivially true. Let us then suppose that $\mathtt{ddi}(\Pi(\mathcal{S})|\mathbb{S}) = \Pi(\mathbb{S})$ and let us fix an arbitrary $M \in \mathfrak{M}_{n \times \ell}$ such that $\operatorname{supp} M = \mathcal{V}$. Then we have $M\Pi = M$. By using lemma 2 we have

$$\begin{aligned} \mathtt{ddi}(M(\mathcal{S})|\mathbb{S}) &= \mathtt{ddi}(M\Pi(\mathcal{S})|\mathbb{S}) \\ &= M\mathtt{ddi}(\Pi(\mathcal{S})|\mathbb{S}) = M\Pi(\mathbb{S}) = M(\mathbb{S}) \end{aligned}$$

and the thesis is proved. ∎

An immediate corollary of this theorem is:

Corollary 2. A set of states $\mathcal{S} \subseteq \mathbb{S}$ is observationally complete for any informationally complete measurement if and only if $\mathtt{ddi}(\mathcal{S}|\mathbb{S}) = \mathbb{S}$.

This is the statement of Theorem 3 in the main text.

If the set of states $\mathbb{S}$ is (hyper)spherical, it is possible to give an explicit characterization of the sets of states with minimum cardinality that are observationally complete for the informationally complete measurements.
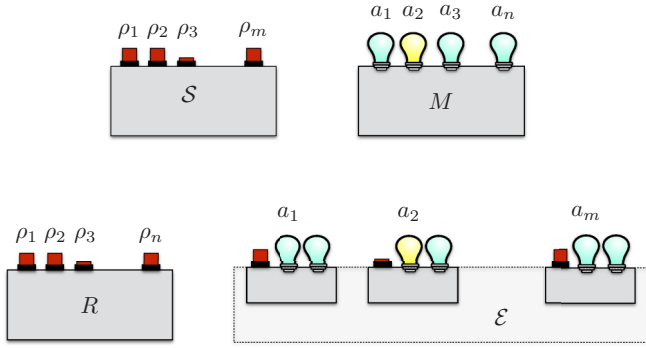
FIG. 12. Top: Setup for the data-driven reconstruction of measurements. On the right is a measurement $\{a_y\}_{y=1}^n$ that can be thought as a box M equipped with n bulbs, one for each effect $a_y \in \mathbb{E}$, $y = 1, 2, \ldots, n$, corresponding to the possible outcomes of the measurement. On the left is a state preparator that can be thought as a box with m possible buttons, each button corresponding to a state $\rho_x \in \mathbb{S}$, $x = 1, \ldots, m$. At each run of the experiment a button is pressed and the outcome of the measurements is recorded. We iterate this procedure many times recording the frequencies p_x^y that are our estimates of the probabilities $a_y(\rho_x)$. Since each state $\rho_x \in \mathbb{S}$ can be regarded as a vector in $\mathbb{R}^\ell$, the whole experiment can be represented by a linear map $M : \mathbb{R}^l \to \mathbb{R}^n$, with $\rho_x \mapsto \mathbf{p}_x$, $p_x^y := a_y(\rho_x)$. Any state in the set $S = \{\rho_x\}_{x=1}^m$ is associated with a point $\mathbf{p}_x \in \mathbb{R}^n$. Bottom: Setup for the data-driven reconstruction of families of states. On the left is a state-preparator R that can prepare one out of a finite set $\{\rho_y\}_{y=1}^n$ of states. We can think of an n-buttons state preparator: whenever we press the button y the state ρ_y is prepared. On the right is a measurement box with m possible buttons: whenever button x (with $x = 1, \ldots, m$) is pressed the two outcomes measurement described by the pair of effects $\{a_x, \bar{a}_x\}$ is applied to the prepared state. Give a state ρ_y the outcome a_x will be obtained with probability $a_x(\rho_y)$ [clearly, $a_x(\rho_y) + \bar{a}_x(\rho_y) = 1$]. We iterate this procedure for all the n states ρ_y and the m measurements $\{a_x, \bar{a}_x\}$, thus recording the frequencies p_x^y, which are our estimate of the probabilities $a_x(\rho_y)$. Since each effect a_x corresponds to a vector in $\mathbb{R}^\ell$, the whole experiment can be represented by a linear map $R : \mathbb{R}^l \to \mathbb{R}^n$, with $a_x \mapsto \mathbf{p}_x$, $p_x^y := a_x(\rho_y)$. Any effect in the set $\mathcal{E} = \{a_x\}_{x=1}^m$ is associated with a point $\mathbf{p}_x \in \mathbb{R}^n$.

Proposition 2. Let $\{\mathbb{S}, \mathbb{E}\}$ be a physical system of linear dimension ℓ and let $\mathbb{S}$ be an $(\ell - 1)$ dimensional (hyper)sphere. Then the following conditions are equivalent:

(1) $\mathcal{S}$ is a regular $(\ell - 1)$ simplex inscribed in $\mathbb{S}$.

(2) $\mathrm{ddi}(\mathcal{S}|\mathbb{S}) = \mathbb{S}$ and $\mathcal{S}$ has minimal cardinality.

Proof. Let us prove each implication separately:

$1 \Rightarrow 2$

Let $\mathcal{A}$ be the regular $(\ell - 1)$ simplex inscribed in $\mathbb{A}$. It is known [46] that $\mathrm{MVEE}(\mathcal{A}) = \mathbb{A}$.

$2 \Rightarrow 1$

Since $\mathbb{S}$ is (hyper)spherical $\mathrm{ddi}(\mathcal{S}|\mathbb{S}) = \mathrm{MVEE}(\mathcal{S}) = \mathbb{S}$. Then $\mathrm{MVEE}(\mathcal{S})$ is the smallest (hyper)sphere which contains $\mathcal{S}$. Clearly the cardinality of $\mathcal{S}$ must be greater then ℓ. On the other hand, as the proof of the previous item shows, the regular simplex has cardinality ℓ and is observationally complete. Therefore $\mathcal{S}$ must be a simplex. We now show that $\mathcal{S}$ is regular. Let us denote with $\mathrm{conv}(\mathcal{S})$ the convex hull of $\mathcal{S}$ and let $r_{\max}$ be the radius of the largest sphere inscribed in $\mathrm{conv}(\mathcal{S})$. Since $\mathrm{MVEE}(\mathcal{S}) \subseteq (\ell - 1)\mathrm{conv}(\mathcal{S})$ [44] we have

$(\ell - 1)r_{\max} \geqslant R$ where R is the radius of $\mathrm{MVEE}(\mathcal{S})$. On the other hand, we have $(\ell - 1)r_{\max} \leqslant R$ from Euler inequality [47]. Therefore $(\ell - 1)r_{\max} = R$ which holds if and only if the simplex is regular. ∎

Let us consider the case when $\ell = 3$ and hence $\mathbb{S}$ is a circle. In this case, any regular polygon $\mathcal{S}$ with n vertices inscribed in $\mathbb{A}$ is $\mathcal{U}$ symmetric, where $\mathcal{U}$ is an orthogonal representation of the dihedral group. Since for $n \geqslant 3$, the only $\mathcal{U}$-symmetric ellipse is the circle, due to Theorem 1 and Corollary 2 any such an $\mathcal{S}$ is observationally complete for any informationally complete measurement.

Let us now consider the case when $\ell = 4$ and hence $\mathbb{S}$ is a sphere. In this case, any platonic solid with n vertices inscribed in $\mathbb{S}$ is $\mathcal{U}$ symmetric, where $\mathcal{U}$ is an orthogonal representation of the tetrahedral (for tetrahedra), octahedral (for octahedra and cubes), or icosahedral (for icosahedra or dodecahedra) group. Since the only $\mathcal{U}$-symmetric ellipsoid is the sphere, due to Theorem 1 and Corollary 2 any such an $\mathcal{S}$ is observationally complete for any informationally complete measurement.

2. Data-driven reconstruction of states

In the protocol of data-driven reconstruction of family of states, Bob is in charge of building the tests (binary-outcome measurements) $\mathcal{E}$ corresponding to the boxes equipped with one button and two light bulbs each in the lower part of Fig. 12. His aim is to enable Alice to correctly infer the family of states R, corresponding to the box with n buttons, up to the equivalence of Theorem 2. In this case, we say that $\mathcal{E}$ is observationally complete for R.

Definition 6 (Observational complete set of effects). Let $\{\mathbb{S}, \mathbb{E}\}$ be a physical system of linear dimension ℓ and $R \in \mathfrak{M}_{n \times \ell}$ be a family of states. A set of effects $\mathcal{E} \subseteq \mathbb{E}$ is *observationally complete for R* if and only if

$$\mathrm{ddi}(R(\mathcal{E})|\mathbb{E}) = R(\mathbb{E}), \tag{D3}$$

where ddi is the data driven inference of states.

Clearly the analogous of Theorem 3 holds.

Theorem 4. Let $\{\mathbb{R}, \mathbb{E}\}$ be a physical system of linear dimension ℓ and let $\mathcal{E} \subseteq \mathbb{E}$ be a set of effects. Let $\mathcal{V}$ be a linear subspace of $\mathbb{R}^\ell$ and let Π denote the projector on $\mathcal{V}$. Then $\mathcal{E}$ is observationally complete for Π if and only of it is observationally complete for any $R \in \mathfrak{M}_{n \times \ell}$ such that $\mathrm{supp} R = \mathcal{V}$, i.e.,

$$\mathrm{ddi}(\Pi(\mathcal{E})|\mathbb{E}) = \Pi(\mathbb{E}) \iff$$
$$\mathrm{ddi}(R(\mathcal{E})|\mathbb{E}) = R(\mathbb{E}) \quad \forall R \text{ s.t. } \mathrm{supp} R = \mathcal{V}. \tag{D4}$$

Proof. The proof of this result is completely analogous to the proof of Theorem 3.

APPENDIX E: TECHNICAL LEMMAS

1. Affine transformations of a strictly affine set

For any $n \in \mathbb{N}$, any $D \in \mathfrak{M}_{n \times \ell}$, and any $\mathbf{d} \in \mathbb{R}^n$, let $\mathcal{F}_{D, \mathbf{d}} : \mathbb{R}^\ell \to \mathbb{R}^n$ denote the affine map such that for any $\mathbf{a} \in \mathbb{R}^\ell$ one has

$$\mathcal{F}_{D, \mathbf{d}}(\mathbf{a}) := D\mathbf{a} + \mathbf{d}.$$

For any set $\mathcal{A} \subseteq \mathbb{R}^\ell$ we adopt the set builder notation

$$\mathcal{F}_{D,\mathbf{d}}(\mathcal{A}) := \{\mathcal{F}_{D,\mathbf{d}}(\mathbf{a}) \mid \mathbf{a} \in \mathcal{A}\}.$$

For any $n \in \mathbb{N}$ and any $\mathcal{A} \subseteq \mathbb{R}^\ell$, let $\mathcal{F}_n(\mathcal{A})$ denote the set of all affine transformations of $\mathcal{A}$ into $\mathbb{R}^n$, that is

$$\mathcal{F}_n(\mathcal{A}) := \{\mathcal{F}_{D,\mathbf{d}}(\mathcal{A}) \mid \forall D \in \mathfrak{M}_{n\times\ell}, \ \forall \mathbf{d} \in \mathbb{R}^n \}.$$

We say that an affine subspace is *strictly* affine if and only if it is not a linear subspace. For any $n \in \mathbb{N}$ and any $\mathcal{X} \in \mathbb{R}^n$, let aff $\mathcal{X}$ denote the affine hull of $\mathcal{X}$. Let $\mathbf{u} \in$ aff $\mathbb{A}$ denote the vector orthogonal to aff $\mathbb{A}$. Without loss of generality we take $\mathbf{u}$ such that $|\mathbf{u}|_2 = 1$.

Lemma 1. For any $n \in \mathbb{N}$ and any $\mathbb{A}$ subset of a strictly affine subspace of $\mathbb{R}^\ell$ one has

$$\mathcal{D}_n(\mathbb{A}) = \mathcal{F}_n(\mathbb{A})$$

[see Eq. (C2) for the definition of $\mathcal{D}_n$].

Proof. Of course $\mathcal{D}_n(\mathbb{A}) \subseteq \mathcal{F}_n(\mathbb{A})$, so we only need to prove the inverse inclusion. For any $D \in \mathfrak{M}_{n\times\ell}$ and $\mathbf{d} \in \mathbb{R}^n$, one has that $D' := D + \mathbf{d}\mathbf{u}^T$ is such that $\mathcal{F}_{D,\mathbf{d}}(a) = D'(a)$ for any $a \in \mathbb{A}$. Hence the thesis follows. ∎

2. Commutativity of the DDI map

Lemma 2. Let $D \in \mathfrak{M}_{n\times\ell}$, and $\mathcal{A} \subseteq \mathbb{R}^\ell$ such that $\mathcal{A} \subseteq \mathrm{supp} D$ one has

$$D(\mathrm{ddi}(\mathcal{A}|\mathbb{A})) = \mathrm{ddi}(D(\mathcal{A})|\mathbb{A}). \tag{E1}$$

Proof. By definition, the left-hand side and right-hand side of Eq. (E1) are given by, respectively,

$$D(\mathrm{ddi}(\mathcal{A}|\mathbb{A})) := D(\mathrm{argmin}_{\mathcal{X}\in\mathfrak{X}}\mathcal{V}(\mathcal{X})),$$
$$\mathrm{ddi}(D(\mathcal{A})|\mathbb{A}) := \mathrm{argmin}_{\mathcal{Y}\in\mathfrak{Y}}\mathcal{V}(\mathcal{Y}),$$

with

$$\mathfrak{X} := \{\mathcal{X} \in \mathcal{D}_\ell(\mathbb{A}) \mid \mathcal{A} \subseteq \mathcal{X} \subseteq \mathrm{aff}\mathcal{A}\},$$
$$\mathfrak{Y} := \{\mathcal{Y} \in \mathcal{D}_n(\mathbb{A}) \mid D(\mathcal{A}) \subseteq \mathcal{Y} \subseteq \mathrm{aff}\, D(\mathcal{A})\}.$$

The map D is bijective from $\mathfrak{X}$ to $\mathfrak{Y}$. This can be easily seen as follows. Since $\mathcal{A} \subseteq \mathrm{supp}\, D$, by definition of $\mathfrak{X}$ for any $\mathcal{X} \in \mathfrak{X}$ one has $\mathcal{X} \subseteq \mathrm{supp}\, D$. Also, by definition of $\mathfrak{Y}$ for any $\mathcal{Y} \in \mathfrak{Y}$ one has $\mathcal{Y} \subseteq \mathrm{supp} D^+$. Moreover, D preserves the ordering induced by function $\mathcal{V}$, that is

$$\mathcal{V}(\mathcal{X}) = \lambda_D \mathcal{V}(\mathcal{L}_D(\mathcal{X})) \quad \forall \mathcal{X} \in \mathfrak{X},$$

for some $\lambda_D > 0$ that only depends on D. Hence the statement follows. ∎

[1] M. Ozawa, Optimal measurements for general quantum systems, Rep. Math. Phys. **18**, 11 (1980).

[2] F. Buscemi, M. Keyl, G. M. D'Ariano, P. Perinotti, and R. F. Werner, Clean positive operator valued measures, J. Math. Phys. **46**, 082109 (2005).

[3] D. Blackwell, Equivalent comparisons of experiments, Ann. Math. Stat. **24**, 265 (1953).

[4] F. Buscemi, Comparison of quantum statistical models: Equivalent conditions for sufficiency, Commun. Math. Phys. **310**, 625 (2012).

[5] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and its Applications* (Springer, Berlin, 1979), Vol. 143.

[6] E. Torgersen, *Comparison of Statistical Experiments*, Encyclopedia of Mathematics and its Applications (Cambridge University Press, Cambridge, 1991).

[7] J. Cohen, J. H. Kempermann, and G. Zbaganu, *Comparisons of Stochastic Matrices with Applications in Information Theory, Statistics, Economics and Population* (Springer Science & Business Media, New York, 1998).

[8] E. H. Lieb and J. Yngvason, The physics and mathematics of the second law of thermodynamics, Phys. Rep. **310**, 1 (1999).

[9] F. Buscemi and N. Datta, Equivalence between divisibility and monotonic decrease of information in classical and quantum stochastic processes, Phys. Rev. A **93**, 012101 (2016).

[10] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, Quantum majorization and a complete set of entropic conditions for quantum thermodynamics, Nat. Commun. **9**, 5352 (2018).

[11] G. Dahl, Matrix majorization, Linear Algebra Appl. **288**, 53 (1999).

[12] E. Shmaya, Comparison of information structures and completely positive maps, J. Phys. A: Math. Gen. **38**, 9717 (2005).

[13] C. E. Shannon, A note on a partial ordering for communication channels, Inform. Control **1**, 390 (1958).

[14] J. Körner and K. Marton, Comparison of two noisy channels, in *Topics in Information Theory*, edited by I. Csiszár and P. Elias, Colloquia Mathamatical Society János Bolyai (North-Holland Publishers, Amsterdam, 1977), pp. 411–423.

[15] F. Liese and K.-J. Miescke, Statistical decision theory, in *Statistical Decision Theory* (Springer, Berlin, 2007), pp. 1–52.

[16] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems* (Cambridge University Press, Cambridge, 2011).

[17] A. Luis and L. L. Sánchez-Soto, Complete Characterization of Arbitrary Quantum Measurement Processes, Phys. Rev. Lett. **83**, 3573 (1999).

[18] G. M. D'Ariano and P. Lo Presti, Quantum Tomography for Measuring Experimentally the Matrix Elements of an Arbitrary Quantum Operation, Phys. Rev. Lett. **86**, 4195 (2001).

[19] G. M. D'Ariano, L. Maccone, and P. Lo Presti, Quantum Calibration of Measurement Instrumentation, Phys. Rev. Lett. **93**, 250407 (2004).

[20] J. Fiurášek, Maximum-likelihood estimation of quantum measurement, Phys. Rev. A **64**, 024102 (2001).

[21] A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, Optimal Quantum Tomography of States, Measurements, and Transformations, Phys. Rev. Lett. **102**, 010404 (2009).

[22] J. Lundeen, A. Feito, H. Coldenstrodt-Ronge, K. Pregnell, C. Silberhorn, T. Ralph, J. Eisert, M. Plenio, and I. Walmsley, Tomography of quantum detectors, Nat. Phys. **5**, 27 (2009).

[23] A. Feito, J. Lundeen, H. Coldenstrodt-Ronge, J. Eisert, M. B. Plenio, and I. A. Walmsley, Measuring measurement: Theory and practice, New J. Phys. **11**, 093038 (2009).

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

[25] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, Found. Phys. **24**, 379 (1994).

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

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

[28] S. Popescu, Nonlocality beyond quantum mechanics, Nat. Phys. **10**, 264 (2014).

[29] G. M. D'Ariano and A. Tosini, Testing axioms for quantum theory on probabilistic toy-theories, Quant. Info. Proc. **9**, 95 (2010).

[30] G. M. D'Ariano, G. Chiribella, and P. Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, Cambridge, 2017).

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

[32] B. Dakic and C. Brukner, Quantum theory and beyond: Is entanglement special? in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, edited by H. Halvorson (Cambridge University Press, Cambridge, 2011), pp. 365–392.

[33] L. Masanes and M. P. Müller, A derivation of quantum theory from physical requirements, New J. Phys. **13**, 063001 (2011).

[34] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Informational derivation of quantum theory, Phys. Rev. A **84**, 012311 (2011).

[35] J. M. Renes, R. Blume-Kohout, A. J. Scott, and C. M. Caves, Symmetric informationally complete quantum measurements, J. Math. Phys. **45**, 2171 (2004).

[36] G. Zauner, Quantendesigns: Grundzuge einer nichtkommutativen Designtheorie, Ph.D. thesis, University of Vienna, 1999, English translation: Quantum designs: Foundations of a noncommutative design theory, also published as [Int. J. Quant. Inf. **9**, 445 (2011)].

[37] C. Fuchs, M. Hoang, and B. Stacey, The SIC question: History and state of play, Axioms **6**, 21 (2017).

[38] C. A. Fuchs and R. Schack, Quantum-Bayesian coherence, Rev. Mod. Phys. **85**, 1693 (2013).

[39] N. D. Mermin, Physics: Qbism puts the scientist back into science, Nat. News **507**, 421 (2014).

[40] C. A. Fuchs, N. D. Mermin, and R. Schack, An introduction to qbism with an application to the locality of quantum mechanics, Am. J. Phys. **82**, 749 (2014).

[41] M. Dall'Arno, S. Brandsen, F. Buscemi, and V. Vedral, Device-Independent Tests of Quantum Measurements, Phys. Rev. Lett. **118**, 250501 (2017).

[42] F. Buscemi and M. Dall'Arno, Device-independent inference of physical devices: Theory and implementation, New J. Phys. **21**, 113029 (2019).

[43] F. John, Extremum problems with inequalities as subsidiary conditions, in *Traces and Emergence of Nonlinear Programming*, edited by G. Giorgi and T. H. Kjeldsen (Springer, Basel, 2014), pp. 197–215.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).

[45] M. J. Todd, *Minimum-Volume Ellipsoids: Theory and Algorithms* (SIAM, Singapore, 2016), Vol. 23.

[46] D. Vandev, A Minimal Volume Ellipsoid around a Simplex, Dokladi na Bulgarskata Akademia na Naukite **45**, 37 (1992).

[47] A. Vince, A simplex contained in a sphere, J. Geom. **89**, 169 (2008).

[48] M. J.Todd and E. A. Yildirim, On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids, Discrete Appl. Math. **155**, 1731 (2007).