

# A de Finetti theorem for quantum causal structures

F. Costa<sup>1,2</sup>, J. Barrett<sup>3</sup>, and S. Shrapnel<sup>4</sup>

<sup>1</sup>Nordita, Stockholm University and KTH Royal Institute of Technology, Hannes Alfvéns väg 12 Stockholm, 106 91, Sweden

<sup>2</sup>School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia

<sup>3</sup>Quantum Group, Department of Computer Science, University of Oxford

<sup>4</sup>ARC Centre for Engineered Quantum Systems, School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia

**What does it mean for a causal structure to be ‘unknown’? Can we even talk about ‘repetitions’ of an experiment without prior knowledge of causal relations? And under what conditions can we say that a set of processes with arbitrary, possibly indefinite, causal structure are independent and identically distributed? Similar questions for classical probabilities, quantum states, and quantum channels are beautifully answered by so-called “de Finetti theorems”, which connect a simple and easy-to-justify condition—symmetry under exchange—with a very particular multipartite structure: a mixture of identical states/channels. Here we extend the result to processes with arbitrary causal structure, including indefinite causal order and multi-time, non-Markovian processes applicable to noisy quantum devices. The result also implies a new class of de Finetti theorems for quantum states subject to a large class of linear constraints, which can be of independent interest.**

## 1 Introduction

The possibility to repeat an experiment is one of the fundamental tenets of the scientific method: after a sufficient number of repetitions, statistical analysis typically enables us to characterise the system or process of interest, to decide between competing hypotheses, and so on. However, a potentially unsettling question underlies this paradigm: what counts as a repetition? If we toss a coin hundreds of thousands of times [1], are we repeating the same experiment or are we

rather performing a set of different experiments? And are we allowed to combine tosses from different coins?

The de Finetti theorem [2–4] offers an elegant answer: it states that a joint probability for a set of random variables that is *a*) invariant under permutations and *b*) the marginal of a probability of an arbitrarily larger set, also invariant under permutations, must take the form

$$P^{(n)}(a^1, \dots, a^n) = \int dq P(q) q(a^1) \cdots q(a^n), \quad (1)$$

for a unique normalised measure  $P(q)$  over the space of single-variable probabilities  $q$ . The striking consequence is that a set of exchangeable variables (i.e., satisfying conditions *a* and *b* above) can always be interpreted as independent and identically distributed (i.i.d.), up to a global uncertainty  $P(q)$  on the single-trial probability  $q$ .

De Finetti’s theorem is particularly satisfying from a Bayesian point of view: one never needs to invoke an ‘unknown’ probability—which would make little sense if probabilities represent degrees of belief—nor needs one introduce any ad-hoc notion of repeatability. As long as they can justify exchangeability, the Bayesian can treat a set of variables *as if* they were multiple trials of the same experiment, with each trial distributed according to the ‘unknown’ probability  $q$ , and where  $P(q)$  represents the prior knowledge about such a probability. Conditioning on a larger and larger set of observed variables allows one to update  $P(q)$ , eventually converging to a particular single-trial  $q^*$ ,  $P(q) \rightarrow \delta(q - q^*)$ . This justifies the idea of ‘discovering’ the unknown probability through repeated trials and gives a principle-based support for Bayesian methods. Apart from the foundational significance, classical de Finetti-

F. Costa: [fabio.costa@su.se](mailto:fabio.costa@su.se)

type theorems have extensive applications in pure and applied probability theory [5–7].

De Finetti’s theorem extends in a direct way to quantum states [8, 9] and channels [10]: an exchangeable multipartite state/channel is always a mixture of product states/channels. Apart from liberating the Bayesian from the uncomfortable notion of ‘unknown state’<sup>1</sup> [11] and grounding the use of Bayesian methods in quantum information [12, 13], these results have wide-ranging applications in many-body physics [14, 15], cryptography [16–18], quantum information [19–23], and quantum foundations [24–27].

However, some natural questions emerging in quantum theory escape known de Finetti results. Consider an experiment that comprises multiple measurements and operations distributed in space and time. Can the causal relations between such operations be *unknown* and discovered through multiple trials? Does it even make sense to *repeat* such an experiment, given that each spacetime event can only happen once?

Within a broader effort to understand the role of causal structure in quantum theory [28–42], a framework has recently emerged in which causal relations need not be fixed in advance [43–45] and can be discovered through experiments. The framework—often dubbed the “process matrix formalism”—also includes scenarios where causal relations are genuinely indefinite [46], with potential applications to quantum information processing [47–51] and fundamental models of quantum gravity [52–55]. Furthermore, the special case of causally ordered, multi-time processes is emerging as a powerful tool to tackle temporal correlations in non-Markovian quantum processes [56–61], an increasingly prominent feature in complex quantum devices [62, 63].

As for ordinary quantum theory, the process matrix formalism makes probabilistic predictions, tacitly assuming that the processes it describes can be repeated arbitrarily many times. Indeed, the paradigm has been employed in several experiments probing processes with definite [64–68] and indefinite [69–76] causal structure. How-

<sup>1</sup>Regardless of one’s ontological view on pure quantum states, it should always be possible to use mixed states to represent incomplete subjective knowledge. In this sense, ‘unknown mixed states’ are as problematic as ‘unknown probability distributions’, calling for a principle-based approach to repeatability and discovery.

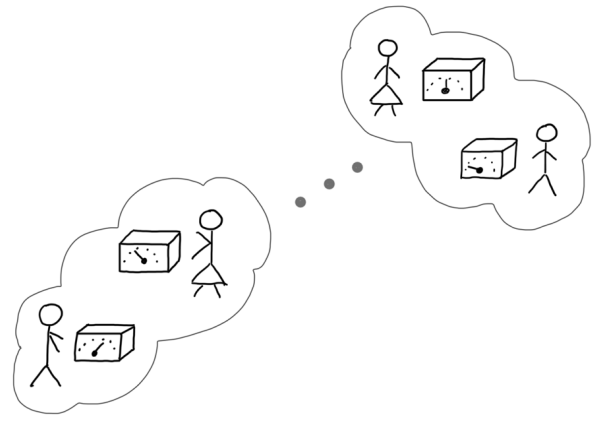


Figure 1: **Process repeatability.** Repetitions of an experiment are modelled as a one-shot scenario, comprising multiple operations that are only performed once. A de Finetti theorem for processes seeks to group these operations such that they represent independent trials under equivalent conditions, with each trial involving multiple operations in arbitrary, possibly unknown or indefinite, causal relations.

ever, a foundational justification for repeatability is missing. In fact, without additional assumptions, even the quantum de Finetti theorem for states does not apply to typical ‘repeated state preparation’ scenarios: If the repetitions are temporally separated, they cannot be modelled as a joint state, as this would constitute a “state over time” [77]. This undermines the applicability of standard statistical analysis and characterisation techniques in the most common quantum experiments.

The goal of this work is to extend the link between exchangeability and repeatability to quantum processes with arbitrary causal structure. To this end, multiple putative trials of an experiment are modelled as a single process, where all measurements and operations are performed only once, Fig. 1. A global process matrix encapsulates prior knowledge about the entire setup, and the task is to establish that an assumption of exchangeability enables the recovery of a decomposition into several repetitions of ‘the same’ process. This requires extending the de Finetti theorem to quantum processes with general causal structures. At first glance, it may seem that any additional causal constraint (such as global causal order or other no-signalling assumptions) necessitates its own de Finetti theorem. This is because, as made clear through the process-state duality, each such assumption is equivalent to a differ-

ent set of linear constraints on quantum states, which, a priori, need not be preserved by the de Finetti representation.

Here we solve the above hurdle by proving a new generalised de Finetti theorem for exchangeable states subject to a broad class of linear constraints. We then specialise the general theorem to show that an exchangeable process subject to a set of no-signalling constraints has a unique representation as a mixture of i.i.d. processes subject to the same constraints. We further consider possible extensions of the result and find that, remarkably, even mild generalisations of the theorem’s hypothesis do not lead to corresponding de Finetti representations.

## 2 The process matrix formalism

The process matrix formalism [43–45] characterises the most general scenario where a quantum system, or multiple quantum systems, can be probed an arbitrary number of times, without prior assumptions regarding spatiotemporal or causal relations between different operations. It relates closely to other frameworks, such as the general boundary formalism [78], higher-order quantum transformations [79–81], multi-time states [82–84], entangled histories [85], and superdensity operators [86].

We call a *site* the abstract location of an operation and we label sites as  $A, B, \dots$ . Concretely, these labels can be understood as spacetime coordinates at which the operations take place, or more general ways to identify the abstract location of the site, which may be delocalised in space and time. The most general operation performed at a site  $A$ , yielding some measurement outcome  $a$ , is a completely positive (CP) and trace non-increasing map [87]  $\mathcal{M}_a : \mathfrak{L}(\mathcal{H}^{A_I}) \rightarrow \mathfrak{L}(\mathcal{H}^{A_O})$ , where  $\mathcal{H}^{A_I}, \mathcal{H}^{A_O}$  are respectively the input and output Hilbert spaces assigned to site  $A$  and  $\mathfrak{L}(\mathcal{H})$  denotes the set of linear operators on  $\mathcal{H}$ . We will always use (a version of) the Choi-Jamiołkowski (CJ) isomorphism to represent CP maps [88, 89]:  $\mathcal{M}_a \mapsto M_a \in \mathfrak{L}(\mathcal{H}^A)$ ,

$$M_a := \left[ \sum_{jk} |j\rangle\langle k| \otimes \mathcal{M}_a(|j\rangle\langle k|) \right]^T, \quad (2)$$

for a chosen orthonormal basis  $\{|j\rangle\}_j$  of  $\mathcal{H}^{A_I}$ , where  $T$  denotes transposition in that basis and

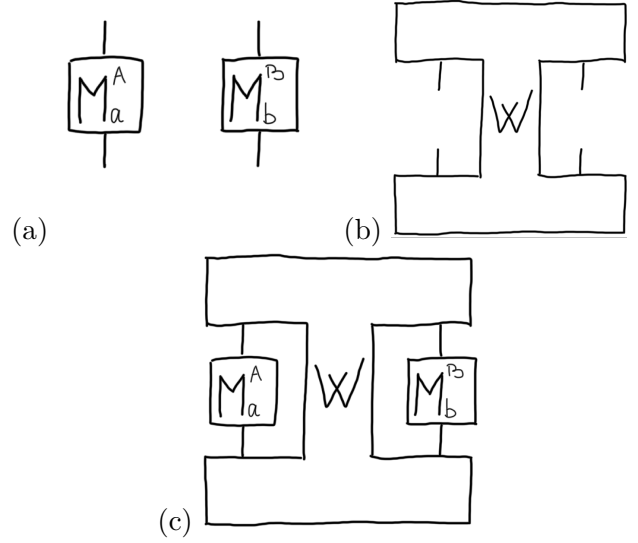


Figure 2: **Operations and processes** (a) Quantum operations, represented as CJ operators  $M_a^A, M_b^B$ , transform an input to an output quantum system—here depicted as wires. The labels  $A, B$  function as generalised coordinates, identifying the operations without necessarily referring to a background causal structure, while  $a, b$  denote measurement outcomes. (b) A process matrix  $W$  represents the most general way to connect operations. (c) Inserting operations into a process, with no open wires left, returns the probability for observing outcomes  $a, b$  through the Born rule for processes, Eq. (4).

we use the short-hand  $\mathcal{H}^A \equiv \mathcal{H}^{A_I} \otimes \mathcal{H}^{A_O}$ . The CP condition translates to the CJ operator being positive semidefinite,  $M_a^A \geq 0$ . For simplicity, we will nominally identify CP maps with their CJ representations.

A deterministic operation  $\bar{M}$ —one with a single measurement outcome that happens with probability one—is represented by a CP and trace preserving (CPTP) map, which in CJ form translates to the condition

$$\text{Tr}_{A_O} \bar{M} = \mathbb{1}^{A_I}, \quad (3)$$

where  $\mathbb{1}^X$  denotes the identity operator on  $\mathcal{H}^X$  (we may skip the superscripts if the context is sufficiently clear). An *instrument* is a collection of CP maps,  $\{M_a\}_a$ ,  $M_a \geq 0$ , that sums to a CPTP map,  $\text{Tr}_{A_O} \sum_a M_a = \mathbb{1}^{A_I}$ . Note that, even when an operation can be regarded as a transformation of a single system, we treat input and output as distinct (although possibly isomorphic) spaces. Furthermore, we assign different Hilbert

spaces to different sites, even though they might represent the same system at different times. The space of operations across all sites spans the tensor product of all input and output spaces.

The probability to obtain a set of outcomes  $a, b, \dots$  at sites  $A, B, \dots$  is given by a generalisation of the Born rule:

$$P(a, b, \dots) = \text{Tr} \left[ \left( M_a^A \otimes M_b^B \dots \right) W \right], \quad (4)$$

where the operator  $W \in \mathfrak{L}(\mathcal{H}^A \otimes \mathcal{H}^B \otimes \dots)$  is the *process matrix*, Fig. 2, which satisfies

$$W \geq 0, \quad (5)$$

$$\text{Tr} \left[ \left( \bar{M}^A \otimes \bar{M}^B \dots \right) W \right] = 1 \quad (6)$$

for all CPTP maps  $\bar{M}^A, \bar{M}^B, \dots$ . These constraints can be derived on abstract grounds by assuming that quantum theory is valid at each site—which can be formalised, e.g., through a non-contextuality assumption [90]—and imposing positivity and normalisation of probabilities. They hold in particular for all ordinary scenarios in quantum mechanics, where all sites can be ordered according to a background time.

The normalisation constraint (6) is equivalent to a set of linear-affine constraints:

$$\text{Tr} W = d^O, \quad (7)$$

$$L(W) = 0, \quad (8)$$

where  $d^O = d^{A_O} d^{B_O} \dots$  is the product of all output-space dimensions and  $L: \mathfrak{L}(\mathcal{H}^A \otimes \mathcal{H}^B \otimes \dots) \rightarrow \mathfrak{L}(\mathcal{H}^A \otimes \mathcal{H}^B \otimes \dots)$  is a linear function whose particular form is not relevant here, see, e.g., Appendix B in Ref. [45].

Additional causal assumptions between sites—specifically, no signalling assumptions—can be enforced by adding linear constraints to the process matrix. For example, imposing no signalling between the sites  $A, B$  of a bipartite process is equivalent to [45]

$$W = (\text{Tr}_{A_O B_O} W) \otimes \mathbb{1}^{A_O B_O}, \quad (9)$$

where (here and in similar expressions) a re-ordering of tensor factors is implied. Similarly, one-way no signalling from  $B$  to  $A$  corresponds to the linear constraints

$$W = (\text{Tr}_{B_O} W) \otimes \mathbb{1}^{B_O}, \quad (10)$$

$$\text{Tr}_{B_I B_O} W = (\text{Tr}_{A_O B_I B_O} W) \otimes \mathbb{1}^{A_O}. \quad (11)$$

Similar constraints characterise one-way no signalling for an arbitrary sequence of causally ordered sites  $A, B, C, \dots$ . Causally ordered processes, also known as quantum channels with memory [91], quantum strategies [92], and quantum combs [93–95], can always be realised as a sequence of channels connecting the sites, possibly with the addition of an auxiliary system (an environment), thus modelling the most general non-Markovian open-system dynamics [56–58]. More complex assumptions, such as an arbitrary partial-order relation between sites, can be obtained by combining non-signalling constraints [37, 40].

In the following, we will need the *reduced process matrix* obtained by removing one or more sites from a larger set [45]. Unlike for states, the reduced process matrix is not always unique: if a site  $B$  has causal influence on (i.e., it can signal to) a site  $A$ , then, by definition of signalling, the reduced process for  $A$  can depend on the operation performed at  $B$ :

$$W_{\bar{M}^B}^A := \text{Tr}_B \left[ \left( \mathbb{1}^A \otimes \bar{M}^B \right) W^{AB} \right], \quad (12)$$

where  $\bar{M}^B = \sum_b M_b^B$  is the CPTP map representing the unconditional transformation at  $B$  if we ignore the outcomes of the instrument  $\{M_b^B\}_b$ .

### 3 Exchangeable processes

As in ordinary quantum mechanics, the process matrix formalism makes probabilistic predictions, through the Born rule in Eq. (4). In order for such predictions to be experimentally meaningful, one typically assumes that an experiment involving a number of sites  $A, B, \dots$  can be repeated an arbitrary number  $n$  of times, where each trial should be modelled by the same process  $W$ . In a consistent probabilistic framework—and in particular, in a Bayesian perspective—it should be possible to combine all trials into a single experiment, and then specify the assumptions that entitle us to view the total experiment as repetitions of individual trials under equivalent conditions.

With this in mind, we will use the labels  $A, B, \dots$  to denote the sites in a single, unspecified trial, while we will use indices  $A^1, \dots, A^n, B^1, \dots, B^n, \dots$  to distinguish the different trials. A priori, an  $n$ -trial scenario with  $m$  sites per trial is described by a single one-shot

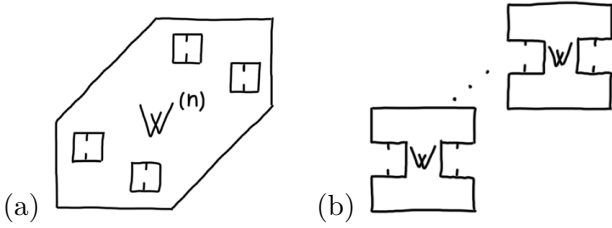


Figure 3: **Exchangeable and i.i.d. processes** (a) A priori,  $n$  repetitions of an experiment constitute a one-shot process  $W^{(n)}$ . (b) The process can be interpreted as  $n$  i.i.d. trials if it is a product of identical processes,  $W^{(n)} = W^{\otimes n}$ . Our result guarantees that exchangeable processes are always mixtures of i.i.d. processes.

scenario with  $N = nm$  sites, with a process matrix

$$W^{(n)} \in \mathfrak{L}(\mathcal{H}_{\text{ST}}^{\otimes n}),$$

Fig. 3, where the single-trial Hilbert space  $\mathcal{H}_{\text{ST}}$  is the tensor product of the input and output spaces associated with all  $m$  single-trial sites:

$$\mathcal{H}_{\text{ST}} := \mathcal{H}^A \otimes \mathcal{H}^B \otimes \dots$$

In the following, the decomposition of an individual trial into sites  $A, B, \dots$  will mostly be irrelevant, so we will effectively treat a single trial as a single site. In line with this, we introduce the short-hand notation  $M^j \equiv M^{A^j} \otimes M^{B^j} \otimes \dots$  to denote a collection of CP maps for trial  $j$ , and simply refer to it as a CP map. With this convention, conditions (5), (6) for a valid  $n$ -trial process matrix can be written as<sup>2</sup>

$$W^{(n)} \geq 0, \quad (13)$$

$$\text{Tr} \left[ \left( \bar{M}^1 \otimes \dots \otimes \bar{M}^n \right) W \right] = 1 \quad (14)$$

for all CPTP maps  $\bar{M}^1, \dots, \bar{M}^n$ .

As in other de Finetti theorems, we need to capture the idea that different trials are equivalent to each other and that, in principle, there is no bound to the number of trials. The first condition is simply modelled by requiring invariance under relabelling of trials. More specifically, given an  $n$ -trial process matrix  $W^{(n)}$ , let us denote by  $|\mu\rangle \in \mathcal{H}_{\text{ST}}$  the states in a chosen orthogonal basis of the single-trial space. We define the

<sup>2</sup>Here and for most of this work, we depart from the common notation in the process matrix literature and drop the superscripts to denote tensor factors. Instead, unless otherwise specified, we assume that tensor factors are always in a reference order, with trials in increasing order from left to right.

action of an  $n$ -element permutation  $\sigma \in S_n$  as

$$U(\sigma) := \sum_{\mu_1 \dots \mu_n} |\mu_1 \dots \mu_n\rangle \langle \mu_{\sigma(1)} \dots \mu_{\sigma(n)}|. \quad (15)$$

This means that all single-trial sites are permuted together:  $A^j \mapsto A^{\sigma(j)}$ ,  $B^j \mapsto B^{\sigma(j)}$ ,  $\dots$  with the same permutation  $\sigma$ . The permuted process matrix is then

$$W_\sigma^{(n)} := U(\sigma)W^{(n)}U^\dagger(\sigma), \quad (16)$$

and we require that it should be equal to the initial process matrix.

The second condition, *extendibility*, means that an  $n$ -trial scenario can be obtained by ‘ignoring’ the last trial in an  $n+1$ -trial scenario:  $\rho^{(n)} = \text{Tr}_{n+1} \rho^{(n+1)}$  for states. For process matrices, there are potentially different candidate definitions, because of the non-uniqueness of the reduced process mentioned above. A simple choice, which is sufficient to deduce a de Finetti representation, is to require that no signalling is possible from the sites in the  $n+1$  trial to all the sites from trial 1 to  $n$ . With this choice, we have the following definition:

**Definition 1.** A sequence of process matrices  $W^{(n)}$ ,  $n \geq 1$ , is called exchangeable if it satisfies

1. **Symmetry:**  $W^{(n)}$  is invariant under permutations of trials:

$$W_\sigma^{(n)} = W^{(n)} \quad \forall \sigma \in S_n; \quad (17)$$

2. **Process extendibility:** For every  $n$  and for every CPTP map  $\bar{M}$ ,

$$W^{(n)} = \text{Tr}_{n+1} \left[ W^{(n+1)} (\mathbf{1}^{\otimes n} \otimes \bar{M}) \right], \quad (18)$$

where the partial trace is over the  $n+1$  factor in  $\mathcal{H}_{\text{ST}}^{\otimes(n+1)}$ .

We say that an  $n$ -trial process matrix is exchangeable if it is part of an exchangeable sequence.

In the definition of extendibility, no-signalling is enforced by requiring that we obtain the same  $W^{(n)}$  for any CPTP map  $\bar{M}$  in the last trial. Together with the symmetry assumption, this implies that no signalling is possible between any (sets of) trials. We explore the consequences of different extendibility conditions in Appendix A.

## 4 Constrained states and processes

Since process matrices are positive semidefinite and have fixed trace, they are always proportional to density operators, up to a constant. This implies that it is possible to treat processes and states in a unified way, which also allows us to leverage known results, such as the quantum de Finetti theorem for states. To this end, it is useful to work with re-normalised process matrices

$$\rho := W / \text{Tr } W = W / d_O. \quad (19)$$

The process matrix normalisation constraints, Eq. (14), are strictly stronger than state normalisation,  $\text{Tr } \rho = 1$ . Therefore, the mapping (19) identifies processes with states that are subject to additional linear constraints. In view of identifying a broader class of constraints, a natural generalisation of Eq. (14) is the following:

**Definition 2.** *Given a set of single-trial operators  $\mathfrak{R} \subset \mathcal{L}(\mathcal{H}_{\text{ST}})$ , and a function  $r : \mathfrak{R} \rightarrow \mathbb{C}$ , we say that an  $n$ -trial operator  $\rho^{(n)}$  satisfies a product expectation constraint if*

$$\text{Tr} \left[ \left( \bigotimes_{j=1}^n R_j \right) \rho^{(n)} \right] = \prod_{j=1}^n r(R_j) \quad (20)$$

for all choices of  $R_1, \dots, R_n \in \mathfrak{R}$ .

The process normalisation constraints, Eq. (14), correspond to

$$\mathfrak{R} = \left\{ R = R^A \otimes R^B \otimes \dots \in \mathcal{L}(\mathcal{H}_{\text{ST}}) \right. \\ \left. \text{s.t. } \text{Tr}_{X_O} R^X = d^{X_O} \mathbb{1}^{X_I}, X = A, B, \dots \right\} \quad (21)$$

and  $r(R) = 1$ , where the conditions in Eq. (21) are the rescaled CJ form of the trace-preserving condition, Eq. (3). More generally, constraints of the form (20) can be interpreted as the assumption that the expectation value of tensor products of certain observables is equal to the product of the expectation values. See Sec. 6 below for a further discussion.

In order to simplify the proof of our main result, and make it directly applicable to a variety of scenarios, it is useful to re-write, and further generalise, constraints of the form (20). We present here the key steps and leave proof details to Appendix B.

The first observation is that the affine component of the constraints—i.e., the term on the

right-hand side of Eq. (20)—can be absorbed in the state constraint  $\text{Tr } \rho^{(n)} = 1$  (which we always assume), while maintaining the product structure of the constraint. This can be stated as follows:

**Lemma 3.** *A sequence of exchangeable states  $\rho^{(n)}$  satisfies product expectation constraints of the form (20), for a set of operator  $\mathfrak{R} \subset \mathcal{L}(\mathcal{H}_{\text{ST}})$  and a function  $r : \mathfrak{R} \rightarrow \mathbb{C}$ , if and only*

$$\text{Tr} \left[ \left( \bigotimes_{j=1}^n \sigma_j \right) \rho^{(n)} \right] = 0 \quad (22)$$

for all

$$\sigma_j := R_j - r(R_j), \quad (23)$$

$R_j \in \mathfrak{R}, j = 1, \dots, n$ .

Next, we observe that a set of linear constraints, such as Eq. (22), can always be rewritten as a single linear constraint defined by an appropriate vector-valued function. Explicitly, we prove in Appendix B.2 that

**Lemma 4.** *Given two inner product spaces  $V^1, V^2$ , a measurable space  $Y$ , and a set of linear functions  $L_y : V^1 \rightarrow V^2, y \in Y$ ,*

$$L_y(v) = 0 \quad \text{a.e. for } y \in Y \quad (24)$$

for every  $v \in V^1$  if and only if, for any strictly positive measure  $q$  over  $Y$  ( $q(y) > 0$  a.e.)

$$\int_Y dy q(y) L_y^\dagger \circ L_y(v) = 0, \quad (25)$$

where “a.e.” stands for “almost everywhere”, meaning that the property in question may not hold in at most a measure-zero set.

As an example of Lemma 4 in action, consider a linear constraint as in Eq. (22), for  $n = 1$ . Defining  $L_R(\rho) := \text{Tr } \sigma \rho$ , where the dependence on  $R$  is given by Eq. (23), the constraint is defined by a (possibly infinite) set of equations:  $L_R(\rho) = 0 \forall R \in \mathfrak{R}$ . However, thanks to Lemma 4, these constraints are equivalent to the single (operator valued) equation

$$\int_{\mathfrak{R}} dR L_R^\dagger \circ L_R(\rho) \\ = \int_{\mathfrak{R}} dR \sigma \text{Tr} [\sigma \rho] = 0. \quad (26)$$

Using the same notation, the constraint in Eq. (22) for  $n > 1$  is given by a set of equations

$L_{R_1} \otimes \cdots \otimes L_{R_n}(\rho^{(n)}) = 0$ , for all combinations of  $R_1, \dots, R_n \in \mathfrak{R}$ . In particular, this implies the constraints  $(L_R)^{\otimes n}(\rho^{(n)}) = 0$ , with the same  $R \in \mathfrak{R}$  for every trial. Using again Lemma 4, this can be expressed with a single equation:

$$\int_{\mathfrak{R}} dR (L_R^\dagger \circ L_R)^{\otimes n}(\rho^{(n)}) = 0. \quad (27)$$

Even though, for generic states, this is a strictly weaker condition than Eq. (22), it will turn out to be sufficient to prove our constrained de Finetti theorem. This motivates us to introduce the following class of *de Finetti-type* constraints:

**Definition 5.** Let  $\mathfrak{H}(\mathcal{H})$  denote the real vector space of self-adjoint operators on a Hilbert space  $\mathcal{H}$ . Given a set of real vector spaces  $\{V^k\}_k$ , a measurable space  $Y$ , a set of linear functions  $L_y^k : \mathfrak{H}(\mathcal{H}_{\text{ST}}) \rightarrow V^k$  defined for all  $y \in Y$ , and a set of strictly positive measures  $q^k$  over  $Y$ , we say that an  $n$ -trial state  $\rho^{(n)}$  satisfies *de Finetti-type constraints* if

$$\int_Y dy q^k(y) (L_y^k)^{\otimes n}(\rho^{(n)}) = 0 \quad \forall k. \quad (28)$$

Note that, using Lemma 4, a set of constraints of the form (28) can always be re-written as a single constraint (eliminating the dependence on  $k$ ). However, for later purposes it is useful to refer to the general form (28).

## 5 Constrained quantum de Finetti theorem

In order to give a unified proof of a constrained de Finetti theorem for processes and states, let us first note that Definition 1 of process exchangeability implies state exchangeability:

**Lemma 6.** *If, for  $n \geq 1$ ,  $W^{(n)}$  define a sequence of exchangeable process matrices, then the normalised operators  $\rho^{(n)} := W^{(n)} / \text{Tr } W^{(n)}$  define a sequence of exchangeable states, i.e., they satisfy the conditions*

1. **Symmetry.**  $\rho^{(n)}$  is symmetric under permutations
2. **State extendibility.** For every  $n$ ,

$$\rho^{(n)} = \text{Tr}_{n+1} \rho^{(n+1)}. \quad (29)$$

To see that this holds, it is sufficient to note that  $\overline{M} = \mathbb{1}/d^{\mathcal{O}}$  is the Choi representation of a CPTP map (the maximally depolarising channel). By substituting this map into the definition of process extendibility, Eq. (18), we obtain state extendibility, Eq. (29), after normalisation.

Note that the result above only works one way: even if we assume that  $\rho^{(n)}$  satisfy the linear constraints for processes, state extendibility is a strictly weaker condition than process extendibility, as discussed in Appendix A.3. However, the mapping from exchangeable processes to exchangeable states is all we need to invoke the standard quantum de Finetti theorem:

**Theorem 7** (De Finetti theorem for quantum states [8, 11]). *If, for  $n \geq 1$ ,  $\rho^{(n)}$  is a sequence of exchangeable states, then there is a unique probability measure  $P$  over the space of single-trial density operators  $\mathfrak{S}$  ( $\rho \geq 0$ ,  $\text{Tr } \rho = 1$ ) such that*

$$\rho^{(n)} = \int_{\mathfrak{S}} d\rho P(\rho) \rho^{\otimes n}. \quad (30)$$

After restoring the normalisation factors, Eq. (30) *almost* gives us a de Finetti representation for processes. The reason this is not quite the desired result is that the integral extends over the whole space of single-trial density operators, which also includes operators that are not valid processes. In order to interpret Eq. (30) as a mixture of i.i.d. processes, we have to show that the probability measure  $P$  has support only on the space of valid single-trial processes, that is, processes that satisfy the constraint (6).

As discussed in the previous section, we prove a more general result that applies to all states subject to de Finetti-type constraints, Eq. (28). We want to show that a set of states  $\rho^{(n)}$  satisfying constraints of the form (28) can be written as a mixture of i.i.d. states subject to the same constraint at the single-trial level. Perhaps surprisingly, a much weaker assumption leads to the desired result: it is sufficient to assume that the constraints hold for two trials. Our central result can then be formulated as follows:

**Theorem 8.** *For a set of indices  $k$ , given a set of real vector spaces  $V^k$ , a measurable space  $Y$ , a set linear functions  $L_y^k : \mathfrak{H}(\mathcal{H}_{\text{ST}}) \rightarrow V^k$  for  $y \in Y$ , and a set of strictly positive measures  $q^k$  over  $Y$ , a state  $\rho^{(n)}$  that is*

1. *exchangeable,*

2. subject to a set of constraints of the form (28) for  $n = 2$ :

$$\int_Y dy q^k(y) L_y^k \otimes L_y^k(\rho^{(2)}) = 0 \quad \forall k \quad (31)$$

has a unique representation of the form

$$\rho^{(n)} = \int_{L_y^k(\rho)=0 \text{ a.e. } y \in Y, \forall k} d\rho P(\rho) \rho^{\otimes n} \quad (32)$$

for some probability measure  $P(\rho)$ , where the notation denotes an integral ranging over single-trial states subject to the constraints  $L_y^k(\rho) = 0 \forall k$  and a.e. for  $y \in Y$ .

Interestingly, this theorem can *almost* be derived as the asymptotic limit of a result in Ref. [96], Theorem 2.3, which considers a finite version of a constrained de Finetti theorem. However, the constraints considered in Ref. [96] are stronger than in our theorem, as they only apply to one factor:  $\mathcal{I}^{\otimes(n-1)} \otimes L(W^{(n)}) = W^{(n-1)} \otimes v$  for some  $v$  in the co-domain of  $L$  and where  $\mathcal{I}$  denotes the identity function. It was left as an open question if a weaker condition, as in our theorem, would still lead to a de Finetti theorem. Our result answers the question in the affirmative for the asymptotic limit of infinite exchangeable sequences.

Let us now prove Theorem 8.

*Proof.* For each vector space  $V^k$ , consider an arbitrary dual vector  $\omega^k$ , that is, an arbitrary linear function  $\omega^k : V^k \rightarrow \mathbb{R}$ . Applying  $\omega^k \otimes \omega^k$  to Eq. (31), we have, for every  $k$ ,

$$\int_Y dy q^k(y) \times (\omega^k \circ L_y^k) \otimes (\omega^k \circ L_y^k)(\rho^{(2)}) = 0. \quad (33)$$

Since  $\rho^{(2)}$  is exchangeable, we can expand it in the de Finetti form, Eq. (30), which substituted into Eq. (33) gives

$$\int_Y dy q^k(y) \int_{\mathfrak{S}} d\rho P(\rho) (\omega^k \circ L_y^k(\rho))^2 = 0. \quad (34)$$

By assumption,  $P(\rho) \geq 0$ ,  $q^k(y) > 0$  a.e., and  $\omega^k \circ L_y^k(\rho) \in \mathbb{R}$ , so the expression (34) is a positive linear combination of non-negative terms  $(\omega^k \circ L_y^k(\rho))^2$ . This means that the expression can only vanish if

$$P(\rho) (\omega^k \circ L_y^k(\rho))^2 = 0 \quad (35)$$

$\forall k$  an a.e. for  $\rho \in \mathfrak{S}$ ,  $y \in Y$ .

This is only possible if, for each  $k$  and for almost all  $\rho$ ,  $y$ , at least one of the two factors vanishes, which means that either

- $P(\rho) = 0$  or
- $\omega^k \circ L_y^k(\rho) = 0$ .

Since this has to hold for every dual vector  $\omega^k$ , we conclude that  $P(\rho) = 0$  a.e. for all  $\rho$  such that  $L_y^k(\rho) \neq 0$ . Therefore, the integral in the de Finetti representation for  $\rho^{(n)}$  can be limited to those states that satisfy  $L_y^k(\rho) = 0 \forall k$  and a.e. for  $y \in Y$ . This yields the desired result, Eq. (32).  $\square$

We will now discuss some consequences of this result.

### General process matrices

Because of Lemmas 3 and 4, Theorem 8 immediately implies that a sequence of exchangeable states subject to product expectation constraints, Eq. (20), for some set  $\mathfrak{R}$  of operators and some function  $r$ , has a unique de Finetti representation over states subject to the same constraints at the single-trial level:

$$\rho^{(n)} = \int_{\text{Tr } R\rho=r(R) \forall R \in \mathfrak{R}} d\rho P(\rho) \rho^{\otimes n}. \quad (36)$$

In particular, by choosing  $\mathfrak{R}$  as in (21) (CPTP maps, up to normalisation),  $r(R) = 1$ , and after the appropriate rescaling of operators, we obtain our original goal:

**Corollary 9.** *An  $n$ -trial process matrix  $W^{(n)}$  is exchangeable if and only if it has a de Finetti representation*

$$W^{(n)} = \int_{\mathfrak{W}} dW P(W) W^{\otimes n}, \quad (37)$$

for a unique probability measure  $P(W)$ , where  $\mathfrak{W}$  denotes the set of single-trial process matrices.

### Processes with no-signalling constraints

Let us now consider  $n$ -trial processes subject to specific no-signalling constraints. These can arise if the relative spatiotemporal location of some sets of sites is fixed and known, or from structural properties of the experimental protocol, such as isolation between components of a device. For example, if  $B^j$  is in the future of  $A^j$  in every trial



$j$ , we can impose that  $W^{(n)}$  permits no signalling from  $B^j$  to  $A^j$ . The goal is to show that we can also restrict the single-trial probability  $P$  to have support on processes that satisfy the same no signalling constraint, i.e., that we can interpret our scenario as consisting of i.i.d. processes with no signalling from  $B$  to  $A$ .

As shown explicitly in Ref. [97], no-signalling from a set of parties  $\mathcal{K}_2$  to a disjoint set of parties  $\mathcal{K}_1$  can be characterised through a linear constraint

$$L^{\mathcal{K}_1 \prec \mathcal{K}_2}(W) = 0, \quad (38)$$

where we refer to Ref. [97] for the explicit form of the linear function  $L^{\mathcal{K}_1 \prec \mathcal{K}_2} : \mathfrak{L}(\mathcal{H}_{\text{ST}}) \rightarrow \mathfrak{L}(\mathcal{H}_{\text{ST}})$ . A given scenario may give rise to multiple no-signalling constraints, corresponding to a set of linear functions  $L^k$ , with each  $k$  labelling no-signalling from a set of sites to another set of sites.

Given an  $n$ -trial process matrix  $W^{(n)}$ , imposing a set of no-signalling conditions for each trial results in the set of constraints

$$\begin{aligned} L^k \otimes \mathcal{I} \cdots \otimes \mathcal{I}(W^{(n)}) &= 0 \quad \forall k, \\ \mathcal{I} \otimes L^k \otimes \cdots \otimes \mathcal{I}(W^{(n)}) &= 0 \quad \forall k, \\ \dots \end{aligned}$$

These constraints can be combined to give

$$L^k \otimes \cdots \otimes L^k(W^{(n)}) = 0 \quad \forall k. \quad (39)$$

This is now a set of de Finetti-type constraints, Eq. (28), so we can apply the constrained de Finetti theorem and obtain the following result:

**Corollary 10.** *An exchangeable  $n$ -trial process matrix  $W^{(n)}$ , subject to single-site no-signalling constraints encoded in a set of linear functions  $L^k$ , has a unique representation as a mixture of i.i.d. single-site process matrices subject to the same no-signalling constraints:*

$$W^{(n)} = \int_{\substack{W \in \mathfrak{W} \\ L^k(W) = 0 \forall k}} dW P(W) W^{\otimes n}. \quad (40)$$

In particular, this result applies to a typical experimental setup where, in each run of the experiment, a sequence of temporally separated measurements is performed. In such a case, we are entitled to impose one-way no-signalling constraints to the global  $n$ -trial process matrix  $W^{(n)}$ . If we can also assume exchangeability, then we

can write the  $n$ -trial process matrix as a mixture of i.i.d. causally ordered processes, recovering a *de Finetti theorem for combs*, which was proven independently [98].

## 6 Frequentist interpretation of the constraints

In line with the original de Finetti theorem, we have presented our result within a Bayesian mind set, where probabilities, states, and processes are representations of an agent's knowledge of (or expectations about) a given scenario, without any a priori frequentist interpretation. The frequentist interpretation emerges for the single-trial probabilities appearing in the de Finetti representation.

However, it can be useful to re-interpret the result in a frequentist approach, as this can give some intuition about when the state constraints we have introduced may arise. For simplicity, we will only consider states, although the discussion directly extends to processes.

In a frequentist view, the  $n$ -trial state  $\rho^{(n)}$ , and any probability we associate to it, describes a scenario where the entire set of trials can be repeated an arbitrary number of times. We refer to a set of  $n$  trials as a *supertrial*. It is of course possible to recover the frequentist interpretation of  $\rho^{(n)}$  by enlarging the Bayesian description to include an unbounded number of supertrials. The fact that each supertrial is represented by the same, 'unknown'  $\rho^{(n)}$  is then recovered by assuming exchangeability of the sequence of supertrials. However, we will stick to a frequentist language for describing the repetitions of supertrials.

For concreteness, we can imagine a source of quantum systems that is turned on every day. Each day, the source produces  $n$  systems, which we collectively describe by some state  $\rho^{(n)}$ , and on which we can perform arbitrary measurements. After repeating the same measurements for an arbitrary number of days, the frequency of any measurement converges to the probability calculated from  $\rho^{(n)}$ . If  $\rho^{(n)}$  is exchangeable (and hence has a de Finetti representation), it means that, within any given day, the source prepares the same state  $\rho$  over and over, but a different  $\rho$  is picked randomly each day according to the probability  $P(\rho)$ , so that the supertrial state is represented by the mixture  $\rho^{(n)} = \int_{\mathfrak{E}} P(\rho) \rho^{\otimes n}$ .

The frequentist language makes it easier to

formulate a context in which constraints on the expectation values of observables, such as in Eq. (20), may arise. Imagine that, each day, we only use the source twice, and each time we measure some chosen observable  $R$ . This means that we are measuring the product observable  $R \otimes R$  over the two-trial state  $\rho^{(2)} = \text{Tr}_{3\dots n} \rho^{(n)}$ . After repeating the experiment over many days, we record the one-trial and two-trial expectation values

$$\langle R \rangle = \text{Tr} \left( R \rho^{(1)} \right), \quad (41)$$

$$\langle R \otimes R \rangle = \text{Tr} \left( R \otimes R \rho^{(2)} \right). \quad (42)$$

Note that, at this point, we do not know what  $\rho^{(n)}$  is—we just assume that there must be some  $\rho^{(n)}$  that reproduces our observed expectation values according to expressions (41) and (42), where  $\rho^{(1)} = \text{Tr}_2 \rho^{(2)}$ .

In this scenario, we say that we have a product expectation constraint if

$$\langle R \otimes R \rangle = \langle R \rangle^2. \quad (43)$$

Indeed, this coincides with the form (20) where the set of observables  $\mathfrak{A}$  contains a single element  $R$ ,  $r(R) = \langle R \rangle$ , and  $n = 2$ .

The constrained de Finetti theorem implies that, having observed the relation (43) (and assuming exchangeability), we can conclude that, at each trial, the source always produces a state  $\rho$  with expectation value

$$\text{Tr} (R \rho) = r(R) = \langle R \rangle. \quad (44)$$

Note that this is not the same as the expectation value (41), which is obtained from one measurement per day repeated over many days, for which we use  $\rho^{(1)} = \int_{\mathfrak{E}} d\rho P(\rho) \rho$ . Eq. (44) means that if, over a *single* day, we measure  $R$  arbitrarily many times (not just two), the average of all observed results will approach  $\langle R \rangle$ . More general product expectation constraints arise in analogous scenarios for a larger number of observables.

As a concrete example, if the single-trial system has two levels, and we take the Pauli observable  $R = Z$ , then the two-trial constraint  $\langle Z \otimes Z \rangle =$

$\langle Z \rangle^2 =: \bar{z}^2$  on an exchangeable state  $\rho^{(n)}$  implies<sup>3</sup>

$$\rho^{(n)} = \int_{x^2+y^2 \leq 1-\bar{z}^2} dx dy P(x, y) \times \frac{1}{2^n} (\mathbb{1} + xX + yY + \bar{z}Z)^{\otimes n}, \quad (45)$$

where  $X$  and  $Y$  are the two remaining Pauli operators. This tells us that the source always prepares a state in the section of the Bloch sphere with  $z$  coordinate equal to  $\bar{z}$ , but the  $x$  and  $y$  coordinates are picked randomly each day according to the distribution  $P(x, y)$ .

## 7 (Non)-extensions of the result

It is natural to ask whether there are other types of constraints, not covered by Theorem 8, that still lead to a constrained de Finetti representation. It is in fact quite instructive that this does not work for some seemingly mild generalisations of our constraints. In rather general terms, we can formulate the question as follows:

*Given a sequence of exchangeable states  $\rho^{(n)}$  and a sequence of functions  $F^{(n)}$  such that, for some values of  $n$ ,  $F^{(n)}(\rho^{(n)}) = 0$ , does  $\rho^{(n)}$  have a de Finetti representation with support on states satisfying  $F^{(1)}(\rho) = 0$ ? In other words, can we write*

$$\rho^{(n)} = \int_{F^{(1)}(\rho)=0} d\rho P(\rho) \rho^{\otimes n} \quad (46)$$

*for some probability measure  $P(\rho)$ ? We explore examples of this type in Appendix C. Here we only summarise the key findings.*

- Instead of constraints that fix the expectation values of some observables through equalities, as in Eq. (20), we can consider *inequalities* instead. As it turns out, inequality constraints such as  $\text{Tr} \left[ \left( \bigotimes_{j=1}^n R_j \right) \rho^{(n)} \right] \geq \prod_{j=1}^n r(R_j)$ , do not imply that the de Finetti representation can be constrained to states with  $\text{Tr} (R_j \rho) \geq r(R_j)$ .
- In Theorem 8, we have seen that it is sufficient to impose the linear constraints on  $\rho^{(2)}$  in order to derive a constrained de Finetti representation. By contrast, imposing the constraints only on  $\rho^{(1)}$ , or on  $\rho^{(n)}$  for any

<sup>3</sup>A similarly constrained state was considered in Ref. [12], although there the constraint emerged from arbitrarily many measurements of  $Z$  within a single supertrial.

odd  $n$ , is generally not sufficient to derive a constrained de Finetti representation.

- However, if  $R$  is positive semidefinite, it is sufficient to impose  $\text{Tr } R\rho^{(1)} = 0$  to derive a de Finetti representation constrained on states with  $\text{Tr } R\rho = 0$ . Since  $R > 0$  can be interpreted as a measurement operator, the implication is that if a measurement outcome cannot occur in a single trial, then the ‘unknown state’ must also give 0 probability for that outcome.
- If, on top of exchangeability, we require  $\rho^{(n)}$  to be invariant under the action of a group acting on the single-trial space, we *cannot* conclude that the corresponding de Finetti representation is constrained to states invariant under the group’s action.

## 8 Conclusions

At the technical level, we have shown that an exchangeable state or process subject to a product of linear constraints, or mixtures thereof, can always be expressed as a mixture of product states or processes, with each factor satisfying the same constraints. In particular, our result implies that exchangeable processes with arbitrary causal structure are mixtures of i.i.d. processes; that exchangeable multi-time, causally ordered processes are mixtures of i.i.d. multi-time, causally ordered processes; and that exchangeable processes with specified no-signalling constraints are mixtures of i.i.d. processes with the same no-signalling constraints.

Our result shows that an exchangeability assumption can ground the notion of repeatability for experiments involving quantum causal structure. The result clarifies under what conditions we are entitled to regard an experiment as multiple repetitions, *under equal conditions*, of a single experiment involving multiple events. This paradigm entitles us to ‘discover an unknown causal structure’ without any ontological commitment regarding quantum states, processes, or causal structure: as long as we can justify the equivalence of different trials under permutations, and the possibility to repeat an experiment indefinitely, we can treat our scenario *as if* it is governed by a ‘real’ underlying process matrix  $W$ ,

which in turn encodes all potentially accessible information about causal relations.

In practice, once exchangeability is established, the distribution  $P(W)$  appearing in the de Finetti representation, Eq. (37), effectively describes prior knowledge of the process of interest. Following a similar argument as for states [12], observing a set of outcomes  $\vec{a}$  in each trial leads to an update of the prior according to Bayes rule:

$$P_{\text{upd}}(W|\vec{a}) = P(W) \frac{P(\vec{a}|W)}{P(\vec{a})}, \quad (47)$$

where  $P(\vec{a}|W)$  is calculated using the Born rule for processes, Eq. (4), and  $P(\vec{a})$  is the marginal outcome probability given the prior:  $P(\vec{a}) = \int dW P(\vec{a}|W)P(W)$ . Crucially, given prior information about the causal relations between the events involved, we can constrain accordingly the prior  $P(W)$ , without additional assumptions apart from exchangeability. This opens the task to formalise the procedure into concrete algorithmic routines, extending existing methods for quantum states and quantum channels [13].

Just like the original de Finetti theorem, our result relies on the unphysical assumption that the experiment can be repeated an arbitrarily large number of times. However, finite de Finetti theorems exist for classical probabilities [99, 100], quantum states [101, 102], and quantum channels [96]. The general idea is that a subsystem of a finite, symmetric state (or channel, process, etc.), should approximate a state (or channel, process, etc.) in the de Finetti form. Formulating a finite version of a de Finetti theorem for processes could provide a concrete starting point to treat practical scenarios where full exchangeability cannot be guaranteed. As noted earlier, finding a finite version of a constrained de Finetti theorem such as Theorem 8 was also posed as an open question in Ref. [96].

Finally, it is interesting that the constraints in our theorem do not include causal separability [45]. This is relevant, for example, in a scenario where all operations are well localised in time, but there is uncertainty about their order: In this case, one can constrain processes to probabilistic mixture of causally ordered ones, which is a special instance of causal separability [44, 97]. However, working from first principles, this only holds for the entire, one-shot process. Even assuming exchangeability, our current result *does not* guarantee that a causally separable process

can be written as a mixture of i.i.d. causally separable ones, and it is a compelling open question whether such an extension holds.

## Acknowledgments

We thank Eric Cavalcanti, Andrew Doherty, Omar Fawzi, Matthew Pusey, and Frank Wilczek for valuable comments and discussions. This work has been supported by the Australian Research Council (ARC) Centre of Excellence for Engineered Quantum Systems (EQUS, CE170100009) and by the John Templeton Foundation through the ID # 62312 grant, as part of the ‘The Quantum Information Structure of Spacetime’ Project (QISS). The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. Nordita is supported in part by NordForsk. We acknowledge the traditional owners of the land on which the University of Queensland is situated, the Turrbal and Jagera people.

## References

- [1] František Bartoš, Alexandra Sarafoglou, Henrik R. Godmann, Amir Sahrani, et al. “Fair coins tend to land on the same side they started: Evidence from 350,757 flips” (2023). [arXiv:2310.04153](https://arxiv.org/abs/2310.04153).
- [2] Bruno De Finetti. “Funzione caratteristica di un fenomeno aleatorio”. In *Atti del Congresso Internazionale dei Matematici: Bologna del 3 al 10 de settembre di 1928*. Pages 179–190. (1929). url: <http://www.brunodefinetti.it/Opere/funzioneCaratteristica.pdf>.
- [3] Bruno de Finetti. “La prévision : ses lois logiques, ses sources subjectives”. *Annales de l’institut Henri Poincaré* **7**, 1–68 (1937). url: <http://eudml.org/doc/79004>.
- [4] Edwin Hewitt and Leonard J. Savage. “Symmetric measures on cartesian products”. *Trans. Am. Math. Soc.* **80**, 470–501 (1955).
- [5] J. F. C. Kingman. “Uses of exchangeability”. *Ann. Probab.* **6**, 183–197 (1978).
- [6] David J. Aldous. “Exchangeability and related topics”. In P. L. Hennequin, editor, *École d’Été de Probabilités de Saint-Flour XIII — 1983*. Pages 1–198. Springer Berlin Heidelberg (1985).
- [7] Raymond J. O’Brien. “Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. Studies in Bayesian Econometrics and Statistics, Vol. 6”. *The Economic Journal* **98**, 883–884 (1988).
- [8] Erling Størmer. “Symmetric states of infinite tensor products of C\*-algebras”. *J. Funct. Anal.* **3**, 48–68 (1969).
- [9] R. L. Hudson and G. R. Moody. “Locally normal symmetric states and an analogue of de Finetti’s theorem”. *Z. Wahrscheinlichkeitstheorie verw Gebiete* **33**, 343–351 (1976).
- [10] Christopher A. Fuchs, Rüdiger Schack, and Petra F. Scudo. “De Finetti representation theorem for quantum-process tomography”. *Phys. Rev. A* **69**, 062305 (2004).
- [11] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. “Unknown quantum states: The quantum de Finetti representation”. *J. Math. Phys.* **43**, 4537–4559 (2002).
- [12] Rüdiger Schack, Todd A. Brun, and Carlton M. Caves. “Quantum bayes rule”. *Phys. Rev. A* **64**, 014305 (2001).
- [13] Christopher Granade, Christopher Ferrie, Ian Hincks, Steven Casagrande, Thomas Alexander, Jonathan Gross, Michal Kononenko, and Yuval Sanders. “QInfer: Statistical inference software for quantum applications”. *Quantum* **1**, 5 (2017).
- [14] M. Fannes, H. Spohn, and A. Verbeure. “Equilibrium states for mean field models”. *J. Math. Phys.* **21**, 355–358 (1980).
- [15] Christian Krumnow, Zoltán Zimborás, and Jens Eisert. “A fermionic de Finetti theorem”. *J. Math. Phys.* **58**, 122204 (2017).
- [16] Renato Renner. “Symmetry of large physical systems implies independence of subsystems”. *Nature Physics* **3**, 645–649 (2007).
- [17] Matthias Christandl, Robert König, and Renato Renner. “Postselection technique for quantum channels with applications to quantum cryptography”. *Phys. Rev. Lett.* **102**, 020504 (2009).
- [18] R. Renner and J. I. Cirac. “de Finetti representation theorem for infinite-dimensional quantum systems and applications to quantum cryptography”. *Phys. Rev. Lett.* **102**, 110504 (2009).

- [19] Fernando G. S. L. Brandão and Martin B. Plenio. “A generalization of quantum stein’s lemma”. *Commun. Math. Phys.* **295**, 791–828 (2010).
- [20] Miguel Navascués, Masaki Owari, and Martin B. Plenio. “Power of symmetric extensions for entanglement detection”. *Phys. Rev. A* **80**, 052306 (2009).
- [21] Fernando G. S. L. Brandão, Matthias Christandl, and Jon Yard. “Faithful squashed entanglement”. *Commun. Math. Phys.* **306**, 805 (2011).
- [22] Fernando G.S.L. Brandão, Matthias Christandl, and Jon Yard. “A quasipolynomial-time algorithm for the quantum separability problem”. In Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing. Page 343–352. STOC ’11 New York, NY, USA (2011). Association for Computing Machinery.
- [23] Fernando G. S. L. Brandão and Aram W. Harrow. “Quantum de Finetti Theorems Under Local Measurements with Applications”. *Commun. Math. Phys.* **353**, 469–506 (2017).
- [24] R. L. Hudson. “Analogues of de Finetti’s theorem and interpretative problems of quantum mechanics”. *Found Phys* **11**, 805–808 (1981).
- [25] Jonathan Barrett and Matthew Leifer. “The de Finetti theorem for test spaces”. *New J. Phys.* **11**, 033024 (2009).
- [26] Matthias Christandl and Ben Toner. “Finite de Finetti theorem for conditional probability distributions describing physical theories”. *J. Math. Phys.* **50**, 042104 (2009).
- [27] Rotem Arnon-Friedman and Renato Renner. “de Finetti reductions for correlations”. *J. Math. Phys.* **56**, 052203 (2015).
- [28] K. B. Laskey. “Quantum Causal Networks” (2007). [arXiv:0710.1200](https://arxiv.org/abs/0710.1200).
- [29] Matthew S Leifer and Robert W Spekkens. “Towards a formulation of quantum theory as a causally neutral theory of bayesian inference”. *Phys. Rev. A* **88**, 052130 (2013).
- [30] Eric G Cavalcanti and Raymond Lal. “On modifications of reichenbach’s principle of common cause in light of bell’s theorem.”. *J. Phys. A: Math. Theor.* **47**, 424018 (2014).
- [31] Tobias Fritz. “Beyond bell’s theorem ii: Scenarios with arbitrary causal structure”. *Comm. Math. Phys.* Pages 1–44 (2015).
- [32] Christopher J. Wood and Robert W. Spekkens. “The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning”. *New J. Phys.* **17**, 033002 (2015).
- [33] Joe Henson, Raymond Lal, and Matthew F Pusey. “Theory-independent limits on correlations from generalized bayesian networks.”. *New J. Phys.* **16**, 113043 (2014).
- [34] Jacques Pienaar and Časlav Brukner. “A graph-separation theorem for quantum causal models.”. *New J. Phys.* **17**, 073020 (2015).
- [35] Rafael Chaves, Christian Majenz, and David Gross. “Information–theoretic implications of quantum causal structures”. *Nat. Commun.* **6** (2015).
- [36] Katja Ried, Megan Agnew, Lydia Vermeyden, Dominik Janzing, Robert W Spekkens, and Kevin J Resch. “A quantum advantage for inferring causal structure”. *Nat. Phys.* **11**, 414–420 (2015).
- [37] Fabio Costa and Sally Shrapnel. “Quantum causal modelling”. *New J. of Phys.* **18**, 063032 (2016).
- [38] Sally Shrapnel and Fabio Costa. “Causation does not explain contextuality”. *Quantum* **2**, 63 (2018).
- [39] John-Mark A. Allen, Jonathan Barrett, Dominic C. Horsman, Ciarán M. Lee, and Robert W. Spekkens. “Quantum common causes and quantum causal models”. *Phys. Rev. X* **7**, 031021 (2017).
- [40] Christina Giarmatzi and Fabio Costa. “A quantum causal discovery algorithm”. *npj Quant. Inf.* **4**, 17 (2018).
- [41] Jonathan Barrett, Robin Lorenz, and Ognjan Oreshkov. “Quantum causal models” (2019). [arXiv:1906.10726v1](https://arxiv.org/abs/1906.10726v1).
- [42] J. C. Pearl and E. G. Cavalcanti. “Classical causal models cannot faithfully explain Bell nonlocality or Kochen-Specker contextuality in arbitrary scenarios”. *Quantum* **5**, 518 (2021).
- [43] O. Oreshkov, F. Costa, and Č. Brukner. “Quantum correlations with no causal order”. *Nat. Commun.* **3**, 1092 (2012).
- [44] Ognjan Oreshkov and Christina Giarmatzi.

- “Causal and causally separable processes”. *New J. of Phys.* **18**, 093020 (2016).
- [45] Mateus Araújo, Cyril Branciard, Fabio Costa, Adrien Feix, Christina Giarmatzi, and Časlav Brukner. “Witnessing causal nonseparability”. *New J. Phys.* **17**, 102001 (2015).
- [46] G. Chiribella, G. M. D’Ariano, P. Perinotti, and B. Valiron. “Quantum computations without definite causal structure”. *Phys. Rev. A* **88**, 022318 (2013).
- [47] M. Araújo, F. Costa, and Č. Brukner. “Computational Advantage from Quantum-Controlled Ordering of Gates”. *Phys. Rev. Lett.* **113**, 250402 (2014).
- [48] Adrien Feix, Mateus Araújo, and Časlav Brukner. “Quantum superposition of the order of parties as a communication resource”. *Phys. Rev. A* **92**, 052326 (2015).
- [49] Philippe Allard Guérin, Adrien Feix, Mateus Araújo, and Časlav Brukner. “Exponential communication complexity advantage from quantum superposition of the direction of communication”. *Phys. Rev. Lett.* **117**, 100502 (2016).
- [50] Ding Jia and Fabio Costa. “Causal order as a resource for quantum communication”. *Phys. Rev. A* **100** (2019).
- [51] Kaumudibikash Goswami and Fabio Costa. “Classical communication through quantum causal structures”. *Phys. Rev. A* **103**, 042606 (2021).
- [52] L. Hardy. “Towards quantum gravity: a framework for probabilistic theories with non-fixed causal structure”. *J. Phys. A: Math. Gen.* **40**, 3081–3099 (2007).
- [53] Magdalena Zych, Fabio Costa, Igor Pikovski, and Časlav Brukner. “Bell’s theorem for temporal order”. *Nat. Commun.* **10**, 3772 (2019).
- [54] Lucien Hardy. “Implementation of the quantum equivalence principle”. In Felix Finster, Domenico Giulini, Johannes Kleiner, and Jürgen Tolksdorf, editors, *Progress and Visions in Quantum Theory in View of Gravity*. Pages 189–220. Springer International Publishing (2020).
- [55] Lachlan Parker and Fabio Costa. “Background Independence and Quantum Causal Structure”. *Quantum* **6**, 865 (2022).
- [56] Kavan Modi. “Operational approach to open dynamics and quantifying initial correlations”. *Scientific Reports* **2**, 581 (2012).
- [57] Simon Milz, Felix A. Pollock, and Kavan Modi. “An introduction to operational quantum dynamics”. *Open Syst. Inf. Dyn.* **24**, 1740016 (2017).
- [58] Felix A. Pollock, César Rodríguez-Rosario, Thomas Frauenheim, Mauro Paternostro, and Kavan Modi. “Operational markov condition for quantum processes”. *Phys. Rev. Lett.* **120**, 040405 (2018).
- [59] Sally Shrapnel, Fabio Costa, and Gerard Milburn. “Quantum markovianity as a supervised learning task”. *Int. J. Quantum Inf.* **16**, 1840010 (2018).
- [60] Christina Giarmatzi and Fabio Costa. “Witnessing quantum memory in non-Markovian processes”. *Quantum* **5**, 440 (2021).
- [61] I. A. Luchnikov, S. V. Vintskevich, H. Ouerdane, and S. N. Filippov. “Simulation complexity of open quantum dynamics: Connection with tensor networks”. *Phys. Rev. Lett.* **122**, 160401 (2019).
- [62] Joshua Morris, Felix A. Pollock, and Kavan Modi. “Quantifying non-markovian memory in a superconducting quantum computer”. *Open Systems & Information Dynamics* **29**, 2250007 (2022).
- [63] Kevin Young, Stephen Bartlett, Robin J. Blume-Kohout, John King Gamble, Daniel Lobser, Peter Maunz, Erik Nielsen, Timothy James Proctor, Melissa Revelle, and Kenneth Michael Rudinger. “Diagnosing and destroying non-markovian noise”. *Technical Report SAND-2020-10396691214*. Sandia National Lab. (SNL-CA) (2020).
- [64] G. A. L. White, C. D. Hill, F. A. Pollock, L. C. L. Hollenberg, and K. Modi. “Demonstration of non-markovian process characterisation and control on a quantum processor”. *Nat Commun* **11**, 6301 (2020).
- [65] K. Goswami, C. Giarmatzi, C. Monterola, S. Shrapnel, J. Romero, and F. Costa. “Experimental characterization of a non-markovian quantum process”. *Phys. Rev. A* **104**, 022432 (2021).
- [66] G.A.L. White, F.A. Pollock, L.C.L. Hollenberg, K. Modi, and C.D. Hill. “Non-

- markovian quantum process tomography”. *PRX Quantum* **3**, 020344 (2022).
- [67] Liang Xiang, Zhiwen Zong, Ze Zhan, Ying Fei, Chongxin Run, Yaozu Wu, Wenyan Jin, Zhilong Jia, Peng Duan, Jianlan Wu, Yi Yin, and Guoping Guo. “Quantify the non-markovian process with intervening projections in a superconducting processor” (2021). [arXiv:2105.03333](#).
- [68] Christina Giarmatzi, Tyler Jones, Alexei Gilchrist, Prasanna Pakkiam, Arkady Fedorov, and Fabio Costa. “Multi-time quantum process tomography of a superconducting qubit” (2023). [arXiv:2308.00750](#).
- [69] Lorenzo M Procopio, Amir Moqanaki, Mateus Araújo, Fabio Costa, Irati A Calafell, Emma G Dowd, Deny R Hamel, Lee A Rozema, Āaslav Brukner, and Philip Walther. “Experimental superposition of orders of quantum gates”. *Nat. Commun.* **6**, 7913 (2015).
- [70] Giulia Rubino, Lee A. Rozema, Adrien Feix, Mateus Araújo, Jonas M. Zeuner, Lorenzo M. Procopio, Āaslav Brukner, and Philip Walther. “Experimental verification of an indefinite causal order”. *Sci. Adv.* **3**, e1602589 (2017).
- [71] Giulia Rubino, Lee Arthur Rozema, Francesco Massa, Mateus Araújo, Magdalena Zych, Āaslav Brukner, and Philip Walther. “Experimental entanglement of temporal orders” (2017). [arXiv:1712.06884](#).
- [72] K. Goswami, C. Giarmatzi, M. Kewming, F. Costa, C. Branciard, J. Romero, and A. G. White. “Indefinite causal order in a quantum switch”. *Phys. Rev. Lett.* **121**, 090503 (2018).
- [73] Yu Guo, Xiao-Min Hu, Zhi-Bo Hou, Huan Cao, Jin-Ming Cui, Bi-Heng Liu, Yun-Feng Huang, Chuan-Feng Li, Guang-Can Guo, and Giulio Chiribella. “Experimental transmission of quantum information using a superposition of causal orders”. *Phys. Rev. Lett.* **124**, 030502 (2020).
- [74] K. Goswami, Y. Cao, G. A. Paz-Silva, J. Romero, and A. G. White. “Increasing communication capacity via superposition of order”. *Phys. Rev. Research* **2**, 033292 (2020).
- [75] Kejin Wei, Nora Tischler, Si-Ran Zhao, Yu-Huai Li, Juan Miguel Arrazola, Yang Liu, Weijun Zhang, Hao Li, Lixing You, Zhen Wang, Yu-Ao Chen, Barry C. Sanders, Qiang Zhang, Geoff J. Pryde, Feihu Xu, and Jian-Wei Pan. “Experimental quantum switching for exponentially superior quantum communication complexity”. *Phys. Rev. Lett.* **122**, 120504 (2019).
- [76] Márcio M. Taddei, Jaime Cariñe, Daniel Martínez, Tania García, Nayda Guerrero, Alastair A. Abbott, Mateus Araújo, Cyril Branciard, Esteban S. Gómez, Stephen P. Walborn, Leandro Aolita, and Gustavo Lima. “Computational advantage from the quantum superposition of multiple temporal orders of photonic gates”. *PRX Quantum* **2**, 010320 (2021).
- [77] Dominic Horsman, Chris Heunen, Matthew F. Pusey, Jonathan Barrett, and Robert W. Spekkens. “Can a quantum state over time resemble a quantum state at a single time?”. *Proc. Math. Phys. Eng. Sci.* **473**, 20170395 (2017).
- [78] Robert Oeckl. “A “general boundary” formulation for quantum mechanics and quantum gravity”. *Phys. Lett. B* **575**, 318–324 (2003).
- [79] G. Chiribella, G. M. D’Ariano, and P. Perinotti. “Transforming quantum operations: Quantum supermaps”. *EPL (Europhysics Letters)* **83**, 30004 (2008).
- [80] Paolo Perinotti. “Causal structures and the classification of higher order quantum computations”. *Pages 103–127*. Springer International Publishing. Cham (2017).
- [81] Alessandro Bisio and Paolo Perinotti. “Theoretical framework for higher-order quantum theory”. *Proc. Math. Phys. Eng. Sci.* **475**, 20180706 (2019).
- [82] Yakir Aharonov, Sandu Popescu, Jeff Tollaksen, and Lev Vaidman. “Multiple-time states and multiple-time measurements in quantum mechanics”. *Phys. Rev. A* **79**, 052110 (2009).
- [83] Ralph Silva, Yelena Guryanova, Nicolas Brunner, Noah Linden, Anthony J. Short, and Sandu Popescu. “Pre- and postselected quantum states: Density matrices, tomography, and kraus operators”. *Phys. Rev. A* **89**, 012121 (2014).
- [84] Ralph Silva, Yelena Guryanova, Anthony J. Short, Paul Skrzypczyk, Nicolas Brunner,

- and Sandu Popescu. “Connecting processes with indefinite causal order and multi-time quantum states”. *New J. Phys.* **19**, 103022 (2017).
- [85] Jordan Cotler and Frank Wilczek. “Entangled histories”. *Physica Scripta* **2016**, 014004 (2016).
- [86] Jordan Cotler, Chao-Ming Jian, Xiao-Liang Qi, and Frank Wilczek. “Superdensity operators for spacetime quantum mechanics”. *J. High Energ. Phys.* **2018**, 93 (2018).
- [87] Teiko Heinosaari and Mário Ziman. “The mathematical language of quantum theory: From uncertainty to entanglement”. *Cambridge University Press*. (2011).
- [88] A. Jamiołkowski. “Linear transformations which preserve trace and positive semidefiniteness of operators”. *Rep. Math. Phys* **3**, 275–278 (1972).
- [89] Man-Duen Choi. “Completely positive linear maps on complex matrices”. *Linear Algebra Appl.* **10**, 285–290 (1975).
- [90] Sally Shrapnel, Fabio Costa, and Gerard Milburn. “Updating the born rule”. *New J. Phys.* **20**, 053010 (2018).
- [91] Dennis Kretschmann and Reinhard F. Werner. “Quantum channels with memory”. *Phys. Rev. A* **72**, 062323 (2005).
- [92] Gus Gutoski and John Watrous. “Toward a general theory of quantum games”. In *Proceedings of 39th ACM STOC*. Pages 565–574. (2006). [arXiv:quant-ph/0611234](https://arxiv.org/abs/quant-ph/0611234).
- [93] G. Chiribella, G. M. D’Ariano, and P. Perinotti. “Quantum circuit architecture”. *Phys. Rev. Lett.* **101**, 060401 (2008).
- [94] G. Chiribella, G. M. D’Ariano, and P. Perinotti. “Theoretical framework for quantum networks”. *Phys. Rev. A* **80**, 022339 (2009).
- [95] A. Bisio, G. Chiribella, G. D’Ariano, and P. Perinotti. “Quantum networks: General theory and applications”. *Acta Phys. Slovaca* **61**, 273–390 (2011).
- [96] Mario Berta, Francesco Borderi, Omar Fawzi, and Volkher B. Scholz. “Semidefinite programming hierarchies for constrained bilinear optimization”. *Mathematical Programming* **194**, 781–829 (2022).
- [97] Julian Wechs, Alastair A Abbott, and Cyril Branciard. “On the definition and characterisation of multipartite causal (non)separability”. *New J. of Phys.* **21**, 013027 (2019).
- [98] Matthew F. Pusey. Private communication (2019).
- [99] Persi Diaconis. “Finite forms of de Finetti’s theorem on exchangeability”. *Synthese* **36**, 271–281 (1977).
- [100] P. Diaconis and D. Freedman. “Finite Exchangeable Sequences”. *Ann. Probab.* **8**, 745 – 764 (1980).
- [101] Robert König and Renato Renner. “A de Finetti representation for finite symmetric quantum states”. *J. Math. Phys.* **46**, 122108 (2005).
- [102] Robert König and Graeme Mitchison. “A most compendious and facile quantum de Finetti theorem”. *J. Math. Phys.* **50**, 012105 (2009).

## A On the definition of extendibility

The definition of process extendibility, Eq. (18), is rather strong: we ask that we recover the same  $W^{(n)}$  from  $W^{(n+1)}$  for every CPTP map  $\bar{M}^{n+1}$  (in particular, together with symmetry, this implies no signalling across different trials). This is a strictly stronger condition than state extendibility, for which we only ask to recover  $W^{(n)}$  for  $\bar{M}^{n+1} = \mathbb{1}/d^O$ —we show this with an explicit example in Sec. A.3. As we have seen, state extendibility is sufficient to prove our main result, Theorem 8, but we can ask whether other definitions work too. We look at two meaningful options, one turning out to be too weak, while the second still being sufficient to prove the de Finetti theorem for processes.

### A.1 Weak extendibility

It is meaningful to consider scenarios where we only impose that the  $n$ -trial process is recovered from the  $n+1$ -trial one for some particular CPTP map.

**Definition 11.** *A sequence of process matrices  $W^{(n)}$ ,  $n \geq 1$ , is weakly extendible if there is a particular CPTP map  $\bar{M}$  such that*

$$W^{(n)} = \text{Tr}_{n+1} \left[ W^{(n+1)} (\mathbb{1}^{A^1 \dots A^n} \otimes \bar{M}) \right]. \quad (48)$$



We say that  $W^{(n)}$  is *weakly exchangeable* if it is symmetric and weakly extendible. The following counterexample shows that weak exchangeability is not sufficient to ensure a de Finetti representation for the processes  $W^{(n)}$ .

**Counterexample.** Consider a scenario with one site  $A$  per trial, with isomorphic input and output spaces  $A_I, A_O$  of finite dimension  $d$ . For  $n$  trials, define the “idle” process

$$W^{A^1 \rightarrow \dots \rightarrow A^n} := \rho \otimes [[\mathbb{1}_d]]^{\otimes n} \otimes \mathbb{1}_d, \quad (49)$$

where  $\mathbb{1}_d$  is the  $d$ -dimensional identity operator and we use the notation

$$[[\mathbb{1}_d]] := \sum_{j,k=1}^d |j\rangle\langle k| \otimes |j\rangle\langle k| \quad (50)$$

to denote the Choi representation of the identity channel (and, as in the rest of this work, we order tensor factors according to  $A^1, \dots, A^n$ ). The idle process is a causally ordered process where the first site,  $A^1$ , receives the state  $\rho$ , and each site is linked to the next one through the identity map, in the order  $A^1 \prec \dots \prec A^n$ .

Now consider the symmetrised process

$$W_{\text{sym}}^{(n)} := \frac{1}{n!} \sum_{\sigma \in S^n} W^{A^{\sigma(1)} \rightarrow \dots \rightarrow A^{\sigma(n)}}, \quad (51)$$

where the sum is over all  $n$ -element permutations. By construction,  $W_{\text{sym}}^{(n)}$  is symmetric under permutations. We can also see that it satisfies weak extendibility for the CPTP map  $\bar{M} = [[\mathbb{1}_d]]$ . Indeed, plugging the identity map into site  $A^{n+1}$  of the permuted idle process,  $W^{A^{\sigma(1)} \rightarrow \dots \rightarrow A^{\sigma(n+1)}}$ , is equivalent to plugging the identity map into site  $A^{\sigma(n+1)}$  of the unpermuted process, Eq. (49), which, for any permutation  $\sigma$ , always gives back the idle process with one less site. Summing over all permutations then yields  $\text{Tr}_{n+1} \left[ W_{\text{sym}}^{(n+1)} (\mathbb{1}^{A^1 \dots A^n} \otimes [[\mathbb{1}_d]]) \right] = W_{\text{sym}}^{(n)}$ , in agreement with the definition of weak extendibility. In summary,  $W_{\text{sym}}^{(n)}$  is a weakly exchangeable process for every  $n$ , but it is clearly not in the de Finetti form<sup>4</sup>, so weak exchangeability is not sufficient to deduce a de Finetti representation.

Note that weak exchangeability for  $\bar{M} = \mathbb{1}^{A_I A_O} / d^{A_O}$  reduces to state exchangeability,

<sup>4</sup>A simple way to prove this is to note that  $W_{\text{sym}}^{(n)}$  allows some signalling between any pair of sites, which is not possible for processes in the de Finetti form.

which we have seen does lead to a de Finetti theorem. Hence, although weak exchangeability does not work in general, it does for specific choices of CPTP  $\bar{M}$  in Eq. (48).

## A.2 Channel extendibility

A process can be seen as a channel from all output spaces to all input spaces, where the Choi representation of the channel coincides with the process matrix. Operationally, this corresponds to restricting each party’s operations to be a measurement followed by an independent state preparation. The state that the parties receive and measure in their input spaces is given by the channel acting on the output spaces. In formulas, a process matrix  $W \in \mathcal{L}(\mathcal{H}^A \otimes \mathcal{H}^B \otimes \dots)$  defines a channel  $\mathcal{W} : \mathcal{L}(\mathcal{H}^{A_O} \otimes \mathcal{H}^{B_O} \otimes \dots) \rightarrow \mathcal{L}(\mathcal{H}^{A_I} \otimes \mathcal{H}^{B_I} \otimes \dots)$  acting as

$$\begin{aligned} \mathcal{W}(\rho^{A_O B_O \dots}) \\ := \text{Tr}_{A_O B_O \dots} \left[ \left( \rho^{A_O B_O \dots} \right)^T W \right]. \end{aligned} \quad (52)$$

In this view, it is meaningful to consider a version of extendibility that follows the definition for channels given by Fuchs, Schack, and Scudo [10]:

**Definition 12.** *A sequence of process matrices  $W^{(n)}$ ,  $n \geq 1$ , is channel extendible if, for every state  $\rho$ ,*

$$\begin{aligned} W^{(n)} \\ = \text{Tr}_{n+1} \left[ W^{(n+1)} (\mathbb{1}^{A^1 \dots A^n} \otimes \mathbb{1}^{A^{n+1}} \otimes \rho^{A^{n+1}}) \right]. \end{aligned} \quad (53)$$

We say that  $W^{(n)}$  is *channel exchangeable* if it is symmetric and channel extendible.

According to this definition, we only require to recover  $W^{(n)}$  when we restrict  $A^{n+1}$  to “trace and re-prepare” operations, where the input is traced out and an arbitrary state is prepared. For example, the process  $W^{A^1 \rightarrow \dots \rightarrow A^n}$  defined in Eq. (49) satisfies channel extendibility (because any state prepared at the last site simply gets traced out), although it does not satisfy symmetry. On the other hand, the symmetrised process  $W_{\text{sym}}^{(n)}$ , Eq. (51) is not channel extendible.

It is easy to see that channel extendibility can replace process extendibility in the de Finetti theorem. Indeed, fixing the re-prepared state  $\rho$  to be the maximally mixed one, channel extendibility

implies that the normalised states  $W^{(n)}/\text{Tr} W^{(n)}$  are state extendible, which is sufficient to prove the constrained de Finetti theorem.

### A.3 State extendibility does not imply process extendibility

We have mentioned in Sec. 5 of the main text that state extendibility does not imply process extendibility. Let us see an explicit example.

**Example.** Consider a scenario with one site per trial, with isomorphic input and output spaces of dimension  $d$ . Consider a particular case of the idle process, Eq. (49), with initial maximally mixed state  $\rho = \mathbb{1}_d/d =: \mathbb{1}^\circ$ , but with reversed causal order  $A^n \prec \dots \prec A^1$ :

$$\begin{aligned} W^{(n)} &= W_{\mathbb{1}^\circ}^{A^n \rightarrow \dots \rightarrow A^1} \\ &:= U(r_n) \mathbb{1}^\circ \otimes [[\mathbb{1}_d]]^{\otimes n} \otimes \mathbb{1}_d U^\dagger(r_n), \end{aligned} \quad (54)$$

where  $r_n$  is the reverse permutation of  $n$  elements,  $r_n(j) := n + 1 - j$ , and  $U(r_n)$  its unitary representation as defined in Eq. (15).

To see that  $W_{\mathbb{1}^\circ}^{A^n \rightarrow \dots \rightarrow A^1}$  is not process extendible (in fact, that is not even channel extendible), consider a trace-replace CPTP map  $\bar{M} = \mathbb{1}_d \otimes \rho$ , where  $\rho \neq \mathbb{1}^\circ$  is an arbitrary non-maximally-mixed state. Plugging  $\bar{M}$  into the site  $A^{n+1}$  (which is the first in the causal order) of  $W^{(n+1)}$  gives an  $n$ -site idle process, in the order  $A^n \prec \dots \prec A^1$ , starting with  $\rho$ :

$$\begin{aligned} &\text{Tr}_{n+1} \left[ W^{(n+1)} (\mathbb{1}^{A^1 \dots A^n} \otimes \bar{M}) \right] \\ &\text{Tr}_{n+1} \left[ W_{\mathbb{1}^\circ}^{A^n \rightarrow \dots \rightarrow A^1} (\mathbb{1}^{A^1 \dots A^n} \otimes \mathbb{1}_d \otimes \rho) \right] \\ &= U(r_n) \rho \otimes [[\mathbb{1}_d]]^{\otimes n} \otimes \mathbb{1}_d U^\dagger(r_n) \\ &=: W_\rho^{A^n \rightarrow \dots \rightarrow A^1}, \end{aligned} \quad (55)$$

which is not the same as  $W^{(n)}$  as defined in Eq. (54).

On the other hand, substituting  $\rho = \mathbb{1}^\circ$  in the above calculation gives back  $W_{\mathbb{1}^\circ}^{A^n \rightarrow \dots \rightarrow A^1} = W^{(n)}$ . This implies that the sequence of states  $\rho^{(n)} := W^{(n)}/d^n$  satisfies state extendibility. Hence we have found a sequence of (causally ordered) processes that is state extendible (after normalisation) but not process (nor channel) extendible.

## B Proofs of lemmas

### B.1 Proof of Lemma 3

To simplify the notation, let us write  $r_j := r(R_j)$ , so we can restate the lemma as

**Lemma.** *A sequence of exchangeable states  $\rho^{(n)}$  satisfies product expectation constraints, of the form*

$$\text{Tr} \left[ \left( \bigotimes_{j=1}^n R_j \right) \rho^{(n)} \right] = \prod_{j=1}^n r_j \quad (56)$$

for a set of operator  $\mathfrak{R} \subset \mathfrak{L}(\mathcal{H}_{\text{ST}})$ , if and only if it satisfies

$$\text{Tr} \left[ \left( \bigotimes_{j=1}^n \sigma_j \right) \rho^{(n)} \right] = 0 \quad (57)$$

for all

$$\sigma_j := R_j - r_j, \quad (58)$$

$R_j \in \mathfrak{R}$ ,  $j = 1, \dots, n$ .

*Proof.* Let us introduce the notation

$$X_j^0 := R_j, \quad X_j^1 := r_j \quad (59)$$

(where, as usual, identity operators are implied in  $r_j \equiv r_j \mathbb{1}$ ).

We can expand the left hand side of Eq. (57) as

$$\begin{aligned} &\text{Tr} \left[ \left( \bigotimes_{j=1}^n \sigma_j \right) \rho^{(n)} \right] \\ &= \sum_{\mu_1 \dots \mu_n = 0,1} \text{Tr} \left[ \left( \bigotimes_{j=1}^n (-1)^{\mu_j} X_j^{\mu_j} \right) \rho^{(n)} \right]. \end{aligned} \quad (60)$$

Let us now assume that Eq. (56) holds for all  $n$ . Using the exchangeability of  $\rho^{(n)}$ , and hence its de Finetti representation, we can see that

$$\text{Tr} \left[ \bigotimes_{j=1}^n X_j^{\mu_j} \rho^{(n)} \right] = \prod_{j=1}^n r_j \quad (61)$$

for all values of  $\mu_1, \dots, \mu_n = 0, 1$ . So Eq. (60) becomes

$$\begin{aligned} &\sum_{\mu_1 \dots \mu_n = 0,1} \text{Tr} \left[ \left( \bigotimes_{j=1}^n (-1)^{\mu_j} X_j^{\mu_j} \right) \rho^{(n)} \right] \\ &= \prod_{j=1}^n \left( \sum_{\mu_j = 0,1} (-1)^{\mu_j} r_j \right) = 0. \end{aligned}$$

This proves that condition (56) implies (57).

For the other direction, we can proceed by induction. For  $n = 1$ , using  $\text{Tr } \rho^{(1)} = 1$  we see directly that

$$\text{Tr} \left[ (R_1 - r_1) \rho^{(1)} \right] = 0 \quad (62)$$

is equivalent to

$$\text{Tr} \left[ R_1 \rho^{(1)} \right] = r_1. \quad (63)$$

Assume now that Eq. (57) holds for some  $n = \bar{n}$  and that Eq. (56) holds for all  $n < \bar{n}$ . This implies

$$\text{Tr} \left[ \bigotimes_{j=1}^{\bar{n}} X_j^{\mu_j} \rho^{(\bar{n})} \right] = \prod_{j=1}^{\bar{n}} r_j \quad (64)$$

for  $\sum_{j=1}^{\bar{n}} \mu_j > 0$ ; that is, for all choices of  $\mu_1, \dots, \mu_{\bar{n}}$  except  $\mu_1 = \dots = \mu_{\bar{n}} = 0$ , which corresponds to the term  $\text{Tr} \left[ \bigotimes_{j=1}^{\bar{n}} R_j \rho^{(\bar{n})} \right]$ . Using again the expansion (60), we have

$$\begin{aligned} \text{Tr} \left[ \left( \bigotimes_{j=1}^{\bar{n}} \sigma_j \right) \rho^{(\bar{n})} \right] \\ = \text{Tr} \left[ \bigotimes_{j=1}^{\bar{n}} R_j \rho^{(\bar{n})} \right] - \prod_{j=1}^{\bar{n}} r_j. \end{aligned} \quad (65)$$

Together with (57), this implies (56) for an arbitrary  $\bar{n}$ , concluding the proof.  $\square$

## B.2 Proof of Lemma 4

Let us re-state the lemma:

**Lemma.** *Given two inner product spaces  $V^1, V^2$ , a measurable space  $Y$ , a set of linear functions  $L_y : V^1 \rightarrow V^2$ ,  $y \in Y$ , and for any  $v \in V^1$*

$$L_y(v) = 0 \quad \text{a.e. for } y \in Y \quad (66)$$

*if and only if, for any strictly positive measure  $q$  over  $Y$  ( $q(y) > 0$  a.e.)*

$$\int_Y dy q(y) L_y^\dagger \circ L_y(v) = 0, \quad (67)$$

*Proof.* If  $L_y^\dagger = 0$  a.e. for  $y \in Y$ , then it is clear that

$$\int_Y dy q(y) L_y^\dagger \circ L_y(v) = 0 \quad (68)$$

for any measure  $q$ . We need to prove that, for any strictly positive  $q$ , Eq. (68) implies  $L_y(v) = 0$  a.e.

for  $y \in Y$ . To this end, it is sufficient to take the inner product of Eq. (68) with  $v$ :

$$\begin{aligned} 0 &= \left\langle v \left| \int_Y dy q(y) L_y^\dagger \circ L_y(v) \right. \right\rangle \\ &= \int_Y dy q(y) \langle L_y(v) | L_y(v) \rangle. \end{aligned} \quad (69)$$

As the integrand is non-negative a.e., and  $q(y)$  is strictly positive, Eq. (69) implies that  $L_y(v) = 0$  a.e. in  $Y$ , which proves the lemma.  $\square$

## C Other classes of constraints

Here we consider some modifications to the hypothesis of Theorem 8, some of which lead to a corresponding de Finetti representation and some of which do not. Below, it will always be assumed that  $\rho^{(n)}$  are exchangeable states for all  $n$ .

### C.1 Inequality constraints

As in the case of product expectation constraints, consider a set of single-trial observables  $\mathfrak{R}$  and a function  $r : \mathfrak{R} \rightarrow \mathbb{R}$ . However, assume we are only given a bound on the expectation values, in the form

$$\text{tr} \left[ \left( \bigotimes_{j=1}^n R_j \right) \rho^{(n)} \right] \geq \prod_{j=1}^n r(R_j). \quad (70)$$

for an exchangeable  $\rho^{(n)}$ . This constraint *does not* imply a representation of the form

$$\rho^{(n)} = \int_{\text{Tr}(R\rho) \geq r(R) \forall R \in \mathfrak{R}} d\rho P(\rho) \rho^{\otimes n}. \quad (71)$$

**Counterexample** As single-trial space, let us take a two-dimensional system with basis vectors  $|0\rangle$  and  $|1\rangle$ . Consider an operator  $R$  such that

$$\langle 0 | R | 0 \rangle = r_0 \geq 0, \quad \langle 1 | R | 1 \rangle = r_1 > r_0. \quad (72)$$

Now consider the exchangeable states

$$\rho^{(n)} := \frac{1}{2} \left( |0\rangle\langle 0|^{\otimes n} + |1\rangle\langle 1|^{\otimes n} \right). \quad (73)$$

These states satisfy

$$\begin{aligned} \text{Tr} \left( R^{\otimes n} \rho^{(n)} \right) &= \frac{1}{2} (r_0^n + r_1^n) \geq r^n, \\ r &:= \frac{r_0 + r_1}{2}. \end{aligned} \quad (74)$$

Eq. (74) is an inequality constraint of the form (70). However,  $\rho^{(n)}$  cannot be decomposed in the form (71): Eq. (73) is already in the de Finetti form, where the probability

$$P(\rho) = \frac{1}{2} [\delta(\rho - |0\rangle\langle 0|) + \delta(\rho - |1\rangle\langle 1|)]$$

does not vanish on  $\rho = |0\rangle\langle 0|$ , and  $\text{Tr}(R|0\rangle\langle 0|) = r_0 < r$ .

## C.2 Single- and odd-trial constraints

### C.2.1 Single- and odd-trial constraints without a constrained de Finetti representation

Let us consider some single-trial operator  $R$  and a real number  $r$ . We can find exchangeable states  $\rho^{(n)}$  that satisfy the single-trial constraint  $\text{Tr} R\rho^{(1)} = r$  but whose de Finetti representation does not have support limited to states with  $\text{Tr} R\rho = r$ .

As a counterexample, we can take again the state in Eq. (73),  $\rho^{(n)} := \frac{1}{2} (|0\rangle\langle 0|^{\otimes n} + |1\rangle\langle 1|^{\otimes n})$ , and as observable the Pauli operator  $R = Z$ . Since  $\rho^{(1)} = \frac{\mathbb{1}}{2}$ , we have  $\langle Z \rangle = 0$ , even though  $\text{Tr} Z\rho = \pm 1$  for the two states appearing in the de Finetti representation,  $\rho = |0\rangle\langle 0|, |1\rangle\langle 1|$ . In fact, for any odd  $n$  we have  $\langle Z^{\otimes n} \rangle = 0 = \langle Z \rangle^n$ , so, in general, imposing a product expectation constraint only on an odd number of trials does not imply a constrained de Finetti representation.

Note also that, since  $\rho^{(n)}$  is symmetric, single-trial constraints can be expressed equivalently as  $\frac{1}{n} \langle R \otimes \mathbb{1} \cdots \otimes \mathbb{1} + \cdots + \mathbb{1} \cdots \otimes \mathbb{1} \otimes R \rangle = r$ , namely as a constraint on the sample mean of the operator  $R$ . The failure of the constrained de Finetti representation tells us (the known fact) that fixing the expectation value of the sample mean of an observable does not fix the expectation value of that observable for the ‘unknown state’.

### C.2.2 Single-trial constraints with a constrained de Finetti representation

Consider a self-adjoint operator  $R \geq r$ ; that is, such that  $R - r$  is positive semidefinite for some given real number  $r$ . In this case, given an exchangeable sequence of states  $\rho^{(n)}$ , it is sufficient to impose the single-trial constraint

$$\text{Tr} R\rho^{(1)} = r \quad (75)$$

to deduce that the de Finetti representation can be restricted to states  $\rho$  such that  $\text{Tr} R\rho = r$ .

*Proof.*  $R \geq r$  implies that, for every state  $\rho$ ,  $\text{Tr} R\rho \geq r$ . Therefore, the de Finetti representation for a single trial gives

$$\text{Tr} R\rho^{(1)} = \int_{\mathfrak{E}} d\rho P(\rho) \text{Tr} R\rho \geq r, \quad (76)$$

with equality only possible if  $P(\rho) = 0$  a.e. for  $\text{Tr} R\rho > r$ . This means that we can write the  $n$ -trial state as

$$\rho^{(n)} = \int_{\substack{\rho \in \mathfrak{E} \\ \text{Tr} R\rho \geq r}} d\rho P(\rho) \rho^{\otimes n}. \quad (77)$$

□

## C.3 States invariant under joint group action – constraints with the wrong sign

Given a group  $G$  and a unitary representation  $U_g$  on the single-trial space,  $g \in G$ , consider exchangeable states  $\rho^{(n)}$  that are invariant under the joint action of  $G$ :

$$U_g^{\otimes n} \rho^{(n)} (U_g^\dagger)^{\otimes n} = \rho^{(n)} \quad \forall g \in G. \quad (78)$$

One may ask whether states of this type can be decomposed as

$$\rho^{(n)} = \int_{U_g \rho U_g^\dagger = \rho} d\rho P(\rho) \rho^{\otimes n}, \quad (79)$$

but this is not necessarily true. A counterexample is given by

$$\rho^{(n)} = \int d\psi |\psi\rangle\langle\psi|^{\otimes n}, \quad (80)$$

where  $d\psi$  is the measure on pure states induced by the Haar measure (i.e., the unique measure invariant under arbitrary unitary transformations).  $\rho^{(n)}$  is exchangeable and invariant under joint local unitaries, however, the probability measure has support over pure states, which are not invariant under arbitrary unitaries.

It is interesting that constraint (78) is *almost* of the form (28) of a de Finetti-type constraint, but it has the wrong sign: defining the linear operators  $L_1^g(\rho) := U_g \rho U_g^\dagger$  and  $L_2^g(\rho) := -\rho$ , we can write the single-trial constraint as  $(L_1^g + L_2^g)(\rho) = U_g \rho^{(1)} U_g^\dagger - \rho^{(1)} = 0$ . However, this choice of functions implies the wrong two-trial constraint: in

$$\sum_{j=1}^2 L_j^g \otimes L_j^g(\rho^{(2)}) = U_g \otimes U_g \rho^{(2)} U_g^\dagger \otimes U_g^\dagger + \rho^{(2)} = 0$$

the sign in front of  $\rho^{(2)}$  is the opposite as compared to (78) for  $n = 2$ .