PAPER • OPEN ACCESS

Updating the Born rule

To cite this article: Sally Shrapnel *et al* 2018 *New J. Phys.* **20** 053010

View the article online for updates and enhancements.

New Journal of Physics

The open access journal at the forefront of physics

PAPER

OPEN ACCESS

Updating the Born rule

Sally Shrapnel[1] ⓘ , Fabio Costa ⓘ and Gerard Milburn

Centre for Engineered Quantum Systems, School of Mathematics and Physics, The University of Queensland, St Lucia, QLD 4072, Australia
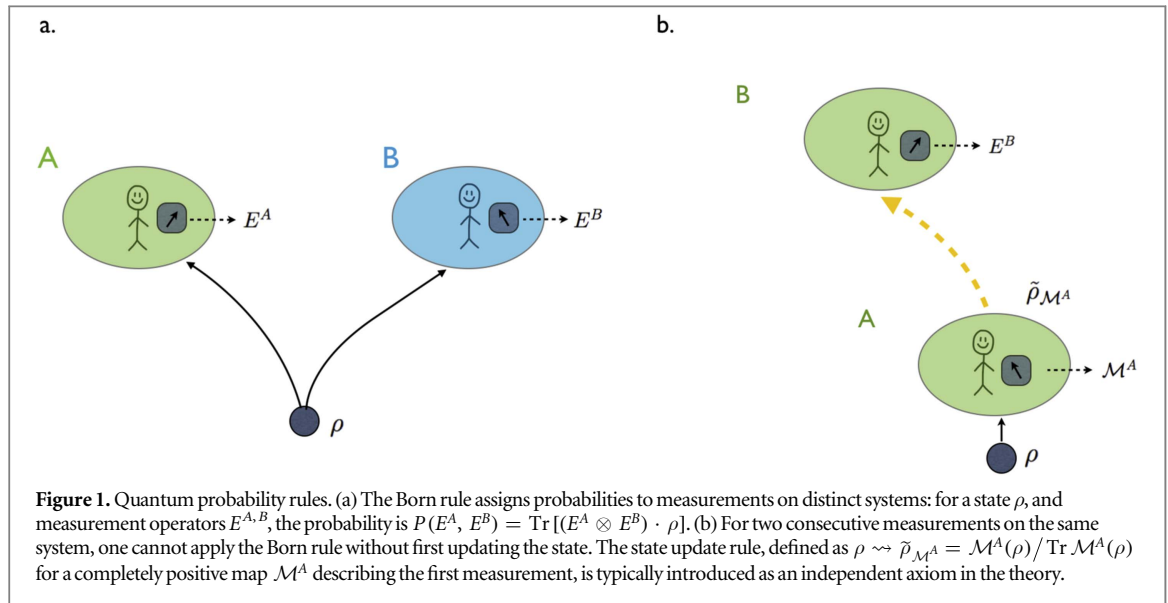[1]   Author to whom any correspondence should be addressed.

E-mail: s.shrapnel@uq.edu.au, f.costa@uq.edu.au and g.milburn@uq.du.au

Abstract

Despite the tremendous empirical success of quantum theory there is still widespread disagreement about what it can tell us about the nature of the world. A central question is whether the theory is about our *knowledge* of reality, or a direct statement about reality itself. Current interpretations of quantum theory, regardless of their stance on this question, regard the Born rule as fundamental and add an independent state update (or 'collapse') rule to describe how quantum states change upon measurement. In this paper we present an alternative perspective and derive a unified probability rule that subsumes *both* the Born rule *and* the collapse rule. We show that this more fundamental probability rule can provide a rigorous foundation for informational, or 'knowledge-based', interpretations of quantum theory. Our result requires an assumption of instrument non-contextuality, a key notion that generalises previous approaches to non-contextuality. Therefore, the framework also permits one to consider non-contextuality in scenarios with arbitrary causal structure.

Introduction

Knowledge-based, or informational, views of quantum theory are popular for a variety of reasons. Perhaps one of the strongest motivations for this perspective comes from the conceptual difficulties that surround quantum state collapse upon measurement. If quantum states are a direct description of reality then this seems to demand that collapse is a nonlinear, stochastic and temporally ill-defined physical process [1–4]. From a 'knowledge' perspective however, collapse is seen merely as a form of information update, no more problematic than classical probabilistic conditioning [5–12]. Whilst compelling, there is an obvious problem with this kind of approach: classical probabilistic conditioning treats multiple consecutive events on a *single* system on exactly the same footing as multiple events on *distinct* systems: joint probabilities are defined in exactly the same way in each case. Thus, classical joint probabilities can be assigned to events in a manner that is independent of the spatio-temporal relationships between those events. In quantum mechanics however, the Born rule does not assign joint probabilities to consecutive events [13], figure 1. This means that knowledge-based interpretations, where one argues that the Born rule is fundamental and the state update rule 'merely a case of probabilistic conditioning', are deeply unsatisfactory. Both rules have to be introduced and justified separately.

In this paper we aim to provide a solution to this problem and breathe new life into the knowledge-based view of quantum theory. We present a proof of a unified quantum probability rule that subsumes *both* the Born rule and the state update rule. This rule is useful in a variety of contexts, from quantum information [14–19] to quantum causal modelling [20–23], and non-Markovian dynamics [24–27]. Dubbed the 'Quantum Process Rule', we prove that one can derive this higher-order, generalised form of the standard quantum probability rule from the structure of quantum operations and a single non-contextuality assumption. This approach is analogous to Gleason's [28] and related [29–32] derivations of the ordinary Born rule. We also show that using this more fundamental approach, where one assigns joint probabilities to arbitrary quantum events, it is possible to derive both the Born rule and the state update rule. A key conceptual advantage is that state update, or 'collapse' need no longer be viewed as an ad hoc ingredient, independent and estranged from the core of the theory.

Figure 1. Quantum probability rules. (a) The Born rule assigns probabilities to measurements on distinct systems: for a state ρ, and measurement operators $E^{A,B}$, the probability is $P(E^A, E^B) = \mathrm{Tr}\,[(E^A \otimes E^B) \cdot \rho]$. (b) For two consecutive measurements on the same system, one cannot apply the Born rule without first updating the state. The state update rule, defined as $\rho \rightsquigarrow \tilde{\rho}_{\mathcal{M}^A} = \mathcal{M}^A(\rho)\,/\,\mathrm{Tr}\,\mathcal{M}^A(\rho)$ for a completely positive map $\mathcal{M}^A$ describing the first measurement, is typically introduced as an independent axiom in the theory.

Results

Measurements and the Born rule

In order to introduce the fewest possible assumptions, we take an explicitly operational perspective. Operational theories can be phrased in terms of *events*, which define the results of *measurements*. Each time a measurement is performed on a system, a number of possible events can be observed. The set of all events that can result from a specific measurement is called a *context*. It is natural when constructing such a theory to assume *measurement non-contextuality* [6, 30]. This means that operationally indistinguishable events should have the same mathematical representation in the theory[2]. Clearly, any probabilistic theory can be formulated in a non-contextual way by appropriate relabelling of the mathematical objects describing events.

In this setting, the *minimal* task of a physical theory is to non-contextually assign probabilities to such measurement events. In essence this is the 'probability rule' of the theory and also defines the relevant state-space. One can represent any such non-contextual probability rule (the Born rule being a prime example) by means of a *frame function*. This is a function that associates a probability to every event, independently of the context to which it belongs, such that probabilities for all events in a given context sum up to one[3]. Crucially, the frame function is *not* a probability distribution over the space of all events, as that would require a normalised measure over the entire space.

As was shown in [29–32], the notion of a frame function can be used to derive the Born rule as the appropriate non-contextual probability rule to apply when one identifies events with the results of a measurement on a quantum system. In this approach, events are identified with *quantum* effects: for a d-level quantum system, the full set of quantum effects is defined as $\mathcal{E}_d := \{E \in \mathcal{L}(\mathcal{H}_d), 0 \leqslant E \leqslant \mathbf{1}\}$, where $\mathcal{L}(\mathcal{H}_d)$ is the space of linear operators on a d-dimensional Hilbert space $\mathcal{H}_d$. Contexts are described by positive-operator-valued measures (POVMs), lists of effect operators $\{E_1, E_2, ...\}$ that sum up to the identity, $\sum_j E_j = \mathbf{1}$. The subscript j labels the list of possible measurement outcomes for a given context.

Assuming measurement non-contextuality here means that the probability of a particular quantum effect (equivalently, measurement outcome) is assumed to be independent of the context (POVM) to which it belongs. Operationally, this means that the probability assigned to a given event does not depend on any extra information regarding how it was achieved.

A frame function for quantum effects is defined as a mapping from the set of all effects to the unit interval:

$$f \colon \mathcal{E}_d \to [0, 1], \tag{1}$$

satisfying

$$\sum_j f(E_j) = 1 \tag{2}$$

for any set $X = \{E_1, E_2, \ldots\}$, $E_j \in \mathcal{E}_d$ such that

$$\sum_j E_j = \mathbf{1}. \tag{3}$$

Using this definition, the task then is to prove that for each frame function, f, there is a unit-trace positive operator ρ such that $f(E_j) = \mathrm{Tr}(\rho E_j)$[4].

The proof in [30] follows three simple steps. First, one proves linearity of the frame function over the field of non-negative rational numbers, then extension to full linearity is obtained by proving continuity of the frame function. Then, as the frame function has been proved to be linear, it can be recast as arising from an inner product. In particular, using the Hilbert–Schmidt inner product on the operator space $\mathcal{L}(\mathcal{H})$, the frame function can be written as $f(E) = \mathrm{Tr}(\rho E)$ for some positive semidefinite, unit-trace operator ρ. This both characterises the Born rule and also defines the density operator as the appropriate object to represent the quantum state.

As we have noted, the above proof does not tell us how to assign probabilities to consecutive events. That is, assuming we know the state of a quantum system prior to measurement, the Born rule alone does not tell us how to update this state following measurement. To remedy this situation, we now wish to provide a similar proof for a probability rule that can subsume *both* the Born rule and the state update rule.

Instruments and the quantum process rule

We consider more general operational primitives than those of [30] and instead consider local regions where one can perform actions that are associated with outcomes. The class of allowed local actions is broad: one can perform measurements, realise transformations, or even add and discard ancillary systems. Such actions can also be associated with local outcomes and we define a particular single case outcome, associated to a given action, as the relevant *event*. The event thus now labels not only the outcome but also any concurrent transformation to the local system.

Just as with effects in the traditional approaches, we assume a minimal operational labelling for transformations: different interactions of the system with an environment, that cannot be distinguished by looking at the system alone, will be assigned the same label.

If we consider a particular run of an experiment there will in general be a collection of such events that occur, one for each local region. One can associate a joint probability to this set of events, and, given enough runs of an experiment, one can empirically verify probability assignments for each possible permutation of events.

Formally, an event in region A is represented by a *completely positive trace-non-increasing* (CP) map $\mathcal{M}^A \colon A_I \to A_O$, where input and output spaces are the spaces of linear operators over input and output Hilbert spaces of the local region, $A_I \equiv \mathcal{L}(\mathcal{H}^{A_I})$, $A_O \equiv \mathcal{L}(\mathcal{H}^{A_O})$ respectively (here identified with the corresponding matrix spaces) [36], see figure 2. We write $L^A := \mathcal{L}(A_I, A_O)$ for the set of linear maps from A_I to A_O. We denote the set of CP maps associated to each region, $\mathrm{CP}^X \subset L^X$.
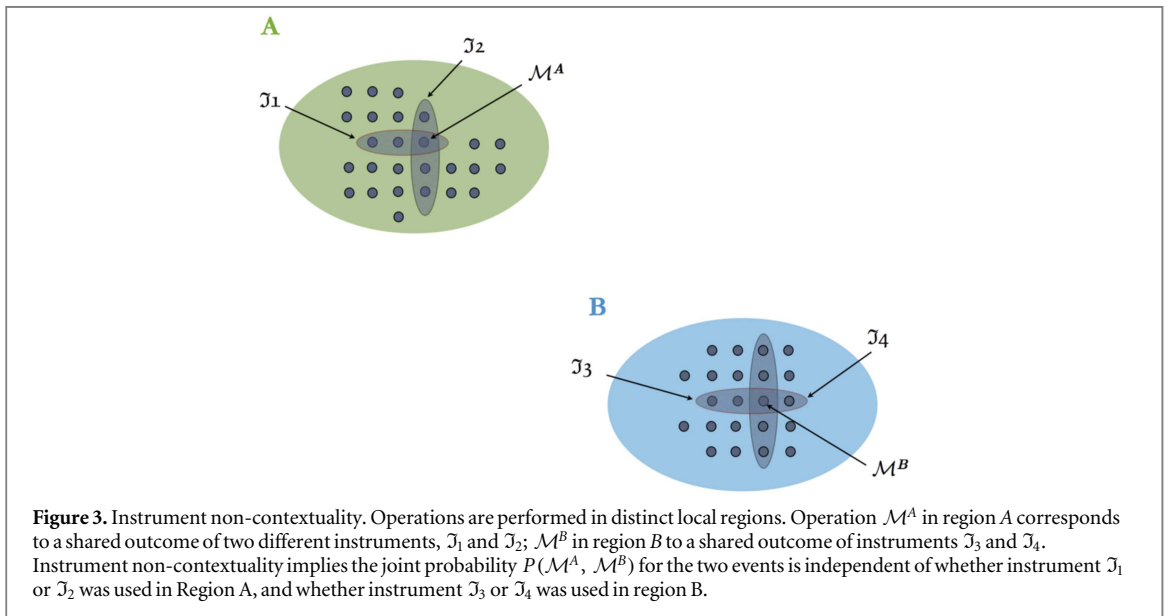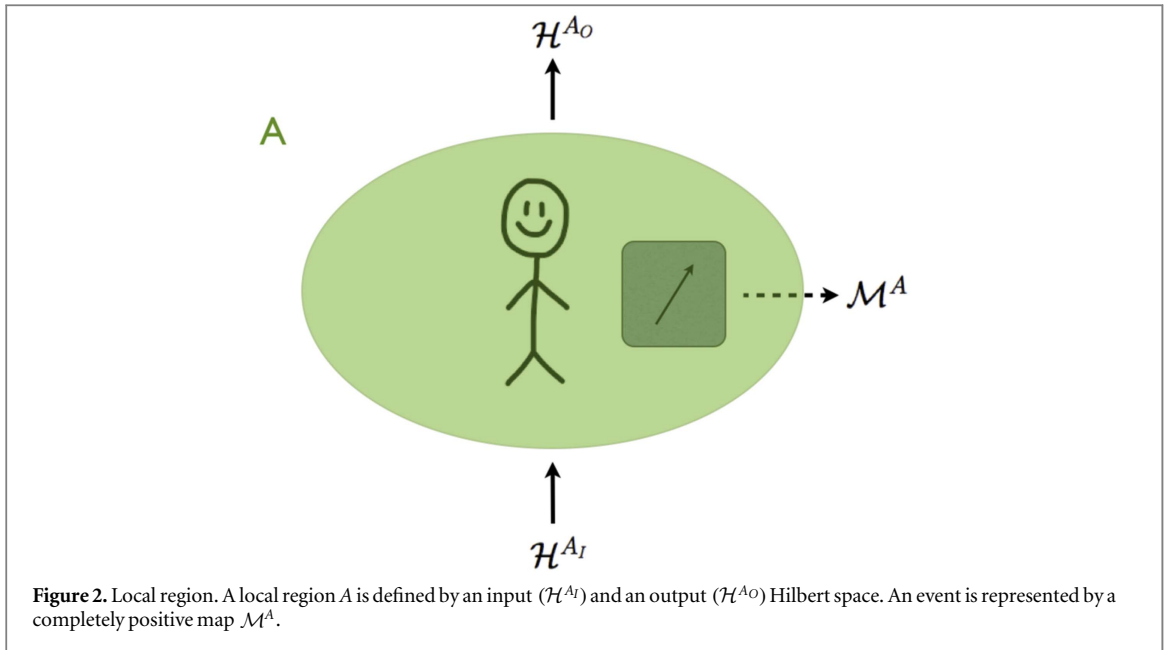
We demand *complete* positivity because in principle it should be possible to perform arbitrary quantum operations in the local region. This includes performing operations on a subsystem that is part of a larger system. Complete positivity means that, for arbitrary dimensions of an ancillary system A', the map $\mathcal{I}^{A'} \otimes \mathcal{M}^A$ transforms positive operators into positive operators, where $\mathcal{I}^{A'}$ is the identity map on A'. Trace non-increasing means that $\mathrm{Tr}\,\mathcal{M}(\rho) \leqslant \mathrm{Tr}\,\rho$ for all operators ρ. A CP map can be decomposed as $\mathcal{M}(\rho) = \sum_{j=1}^m K_j \rho K_j^\dagger$, where the Kraus operators $K_j \colon \mathcal{H}^{A_I} \to \mathcal{H}^{A_O}$, $j = 1, \ldots, m$, satisfy $\sum_{j=1}^m K_j^\dagger K_j \leqslant \mathbf{1}$ for a trace non-increasing map [37, 38].

The context for each set of CP maps is now no longer a POVM but rather a *quantum instrument* [39]. An instrument thus represents the collection of all possible events that can be observed given a specific choice of local action. Given a local region A, an instrument is formally defined as a set $\mathfrak{I}^A$ of CP maps that sum up to a completely positive trace-preserving (CPTP) map:

$$\mathrm{Tr}\left[\sum_{\mathcal{M}^A \in \mathfrak{I}^A} \mathcal{M}^A(\rho) \right] = \mathrm{Tr}(\rho). \tag{4}$$

We are now in a position to define the relevant frame function and derive the appropriate probability rule for this scenario. Just as the Born rule tells us how to calculate the probability of a particular outcome given the relevant measurement operator, the Quantum Process Rule should tell us how to assign a joint probability to each possible collection of local events given the relevant instruments. We assume 'instrument' non-contextuality, rather than 'measurement' non-contextuality. That is, the joint probability for a set of events, one for each region, is independent of the particular context (set of instruments) to which they belong, see figure 3.

---

[4] Henceforth we omit the subscript j for notational convenience.

Figure 2. Local region. A local region *A* is defined by an input ($\mathcal{H}^{A_I}$) and an output ($\mathcal{H}^{A_O}$) Hilbert space. An event is represented by a completely positive map $\mathcal{M}^A$.

Figure 3. Instrument non-contextuality. Operations are performed in distinct local regions. Operation $\mathcal{M}^A$ in region *A* corresponds to a shared outcome of two different instruments, $\mathfrak{I}_1$ and $\mathfrak{I}_2$; $\mathcal{M}^B$ in region *B* to a shared outcome of instruments $\mathfrak{I}_3$ and $\mathfrak{I}_4$. Instrument non-contextuality implies the joint probability $P(\mathcal{M}^A, \mathcal{M}^B)$ for the two events is independent of whether instrument $\mathfrak{I}_1$ or $\mathfrak{I}_2$ was used in Region A, and whether instrument $\mathfrak{I}_3$ or $\mathfrak{I}_4$ was used in region B.

As for [30], the non-contextuality assumption is formalised by requiring that probabilities are given by a frame function. Each 'frame' is now a collection of instruments, one per region, rather than a single POVM.

Definition 1. A frame function, *f*, for a set of local non-intersecting regions $X = A, B, C, ...,$ is defined by:

1. *f* is a function from the Cartesian product of the set of CP maps associated to each region, $\mathrm{CP}^X \subset L^X$, to the unit interval:

$$f: \mathrm{CP}^A \times \mathrm{CP}^B \times \mathrm{CP}^C ... \to [0, 1] \tag{5}$$

2. *f* is normalised for all sets of CP maps, $\mathcal{M}^X$, that form instruments $\mathfrak{I}^X$,

$$\sum_{\substack{\mathcal{M}^A \in \mathfrak{I}^A \\ \mathcal{M}^B \in \mathfrak{I}^B \\ \mathcal{M}^C \in \mathfrak{I}^C \\ ...}} f(\mathcal{M}^A, \mathcal{M}^B, \mathcal{M}^C, ...) = 1 \tag{6}$$

We now show that this definition is sufficient to derive the new probability rule. As in [30] we first prove linearity of the frame function.

Theorem 1. *The frame function f is a convex-multilinear functional on* $\mathrm{CP}^A \times \mathrm{CP}^B \times \mathrm{CP}^C \times ...$

By convex-multilinear we mean:

$$f[p\mathcal{M}_1^A + (1-p)\mathcal{M}_2^A, \mathcal{M}^B, \mathcal{M}^C, ...]$$
$$= pf[\mathcal{M}_1^A, \mathcal{M}^B, \mathcal{M}^C, ...] + (1-p)f[\mathcal{M}_2^A, \mathcal{M}^B, \mathcal{M}^C, ...]$$
$$(0 \leqslant p \leqslant 1)$$

and similarly for all other regions $B, C, ...$

Proof. We fix instruments at all regions, except for region A, to be instruments with a single CPTP map each: $\overline{\mathcal{M}}^B, \overline{\mathcal{M}}^C, ...$

Consider two instruments applied in region A:

$$\mathfrak{I}_1^A = \{\mathcal{M}_1^A, \quad \mathcal{M}_2^A, \quad \mathcal{M}_3^A\},$$
$$\mathfrak{I}_2^A = \{\mathcal{M}_1^A + \mathcal{M}_2^A, \quad \mathcal{M}_3^A\}.$$

The frame function constraints imply:

$$f[\mathcal{M}_1^A, \overline{\mathcal{M}}^B, ...] + f[\mathcal{M}_2^A, \overline{\mathcal{M}}^B, ..] + f[\mathcal{M}_3^A, \overline{\mathcal{M}}^B, ..] = 1$$
$$f[(\mathcal{M}_1^A + \mathcal{M}_2^A), \overline{\mathcal{M}}^B, ...] + f[\mathcal{M}_3^A, \overline{\mathcal{M}}^B, ...] = 1.$$

Therefore

$$f[\mathcal{M}_1^A, \overline{\mathcal{M}}^B, ...] + f[\mathcal{M}_2^A, \overline{\mathcal{M}}^B, ..]$$
$$= f[(\mathcal{M}_1^A + \mathcal{M}_2^A), \overline{\mathcal{M}}^B, ...]$$

and thus we have additivity for CP, trace non-increasing maps.

We next prove homogeneity of the frame function for the rational numbers between 0 and 1. Take two integers $1 \leqslant n \leqslant m \in \mathbb{N}$, and a CP, trace-non-increasing map $\mathcal{M}$. By converting multiplications by integers into sums, from additivity of the frame function we have:

$$\frac{n}{m}f(\mathcal{M}) = \frac{n}{m}f\left(\sum_{j=1}^{m}\frac{\mathcal{M}}{m}\right)$$
$$= \frac{n}{m}\sum_{j=1}^{m}f\left(\frac{\mathcal{M}}{m}\right) = \sum_{i=1}^{n}f\left(\frac{\mathcal{M}}{m}\right)$$
$$= f\left(\sum_{i=1}^{n}\frac{1}{m}\mathcal{M}\right) = f\left(\frac{n}{m}\mathcal{M}\right). \tag{7}$$

Therefore, having additivity and homogeneity, we have proved the convex linearity of *f* for rational numbers between 0 and 1.

Convex linearity of the frame function on the real interval [0, 1] can be established using the 'squeeze theorem' of elementary calculus [40]. Define two sequences of positive rationals, $\{a_n\}$ increasing and $\{b_n\}$ decreasing with $a_n < b_n \leqslant 1$, that converge to the same real number c. Then, for any CP map $\mathcal{M}^A$, the map $\mathcal{N}_n^A := (c - a_n)\mathcal{M}^A$ is also CP. Thus, fixing all maps in other regions to be CPTP, we have

$$f(c\mathcal{M}^A, ...)$$
$$= f(a_n\mathcal{M}^A, ...) + f(\mathcal{N}_n^A, ...) \geqslant f(a_n\mathcal{M}^A, ...).$$

Similarly, we have that $f(c\mathcal{M}^A, ...) \leqslant f(b_n\mathcal{M}^A, ...)$. This implies

$$a_n f(\mathcal{M}^A, ...) \leqslant f(c\mathcal{M}^A, ...) \leqslant b_n f(\mathcal{M}^A, ...). \tag{8}$$

Because $a_n f(\mathcal{M}^A, ...)$ and $b_n f(\mathcal{M}^A, ...)$ both converge to $cf(\mathcal{M}^A, ...)$, equation (8) implies

$$f(c\mathcal{M}^A, ...) = cf(\mathcal{M}^A, ...) \tag{9}$$

by the 'squeeze theorem'.

We have thus proved that *f* is linear on CP^A and, with similar steps, linearity can be proved for $\mathrm{CP}^B, \mathrm{CP}^C, ...$ which concludes the proof. □

Just as in ordinary quantum mechanics a *state* is defined as a linear functional over effects (POVM elements), we can define a multilinear functional over sets of events (CP maps) as a *process*, in accordance with the terminology of [20, 22, 41–48].

We next use the fact that a linear functional can be expressed by means of an inner product. This enables us to derive a new probability rule using our frame function, and also gives the appropriate form for the matrix representation of a process.

First consider that because each convex space CP^X contains a basis of the entire linear space L^X, $X = A, B, ...$, the frame function f can be extended by linearity to a multilinear function on the spaces $L^A, L^B, L^C, ...$ Again by linear extension, this defines a unique linear function on the product space $L^A \otimes L^B \otimes L^C \otimes ...$

Next, it is easy to show that a natural inner product between any two linear maps $\mathcal{M}^A, \mathcal{N}^A \in L^A$ is defined as follows (see the appendix for details):

$$(\mathcal{M}^A, \mathcal{N}^A) := \sum_\mu \mathrm{Tr}\,[\mathcal{M}^A(\tau_\mu)^\dagger \mathcal{N}^A(\tau_\mu)], \tag{10}$$

where $\{\tau_\mu\}_{\mu=0}^{d^2-1}$ is a Hilbert–Schmidt basis for the d-dimensional input space: $\tau_\mu \in \mathcal{L}(\mathcal{H}^{A_I})$, $\tau_\mu = \tau_\mu^\dagger$, $\mathrm{Tr}\,(\tau_\mu \tau_\nu) = \delta_{\mu\nu}$.

One can also represent this inner product in a more convenient (and familiar) form by representing the CP maps associated to each region as Choi–Jamiolkowski (CJ) matrices [49, 50]. Recall, a CP map associated to a region A, where input and output spaces are the spaces of linear operators over input and output Hilbert spaces, $A_I \equiv \mathcal{L}(\mathcal{H}^{A_I})$, $A_O \equiv \mathcal{L}(\mathcal{H}^{A_O})$, respectively, can be represented as a matrix[5]:

$$M^A = \sum_{j,l} |l\rangle \langle j|^{A_I} \otimes [\mathcal{M}(|j\rangle \langle l|)^{A_O}]^T, \tag{11}$$

where ^{T} denotes transposition in a chosen basis and $\{|j\rangle\}_{j=1}^{d_{A_I}}$ is an orthonormal basis in $\mathcal{H}^{A_I}$. We show in the appendix that the inner product (10) can be expressed as

$$(\mathcal{M}^A, \mathcal{N}^A) = \mathrm{Tr}\,(M^{A\dagger} N^A) \tag{12}$$

and it is independent of the choice of Hilbert–Schmidt basis.

This inner product determines an isomorphism between elements of $L^A \otimes L^B \otimes L^C \otimes ...$ and linear functionals on the same space. We can thus define a trace rule that allows one to determine the joint probability for a set of CP maps, one for each region:

$$\begin{aligned} f(\mathcal{M}^A, \mathcal{M}^B, ...) &= (\mathcal{W}_f, \mathcal{M}^A \otimes \mathcal{M}^B \otimes ...) \\ &= \mathrm{Tr}\,[(M^A \otimes M^B \otimes ...) \cdot W_f^{AB\cdots}], \end{aligned} \tag{13}$$

where $\mathcal{W}_f \in L^A \otimes L^B \otimes L^C \otimes ...$ is the linear map that uniquely defines f and $W_f^{AB\cdots}$ is its CJ representation, called the *process matrix*. (In the following, we will drop the subscript f.)

Positivity and normalisation of the frame function, equations (5) and (6) respectively, impose constraints on the operators W that define valid processes. The set of process matrices, together with the expression (13) for the frame function, defines the Quantum Process Rule. As discussed in [20], the set of matrices characterised by positivity, equation (5), is further restricted if one assumes that local operations can be extended to act on additional multipartite quantum states shared among the regions. The overall result can be summarised as follows:

Theorem 2. *Given a set of regions* $X = A, B, ...$ *where arbitrary quantum operations can be performed, any instrument non-contextual probability assignment, expressed through a frame function as per definition 1, must be given by the Quantum Process Rule, equation (13), where the process matrix W satisfies the conditions*

$$\mathrm{Tr}[(\overline{M}^A \otimes \overline{M}^B \otimes ...) \cdot W^{AB\cdots}] = 1 \tag{14}$$

$$\forall\ \overline{M}^X \text{ s.t. } \mathrm{Tr}_{X_O} \overline{M}^X = \mathbb{1}^{X_I} \tag{15}$$

and

$$\mathrm{Tr}\,[(M^A \otimes M^B \otimes ...) \cdot W^{AB\cdots}] \geqslant 0$$
$$\forall\ M^X \geqslant 0. \tag{16}$$

If one additionally assumes that each operation in region X can be extended to act on an additional input space X_I', with an arbitary multipartite state $\rho^{A_I' B_I' \cdots}$ shared across the regions, then the process matrix must be positive semidefinite, $W \geqslant 0$, a strictly stronger condition than equation (16).

---

[5] This definition aligns with the convention in [20]. Other definitions, differing by a transpose or partial transpose, do not change the representation of the inner product.

Property (15) is the CJ representation of the trace-preserving condition; therefore, the normalisation constraint says that CPTP maps can be performed with unit probability. The resulting constraint on process matrices differs from the analogous one for density matrices, $\mathrm{Tr}\,\rho = 1$. In addition to an analogous affine constraint, $\mathrm{Tr}\,W = d_O$ with d_O the product of all output dimensions, W has to satisfy further linear constraints. We refer to appendix B of [42] for an explicit characterisation of such constraints.

Recovering the state update and Born rule

Let us recapitulate the rationale so far: it was shown in [30] that if we accept the structure of quantum measurements, we can identify quantum probabilities as the most general non-contextual probability assignments. Whereas this approach only considers a single measurement/event—or at most measurements of separate quantum systems—in the quantum process approach outlined above we derive a general rule to assign joint probabilities to an arbitrary number of events. The ordinary Born rule is thus recovered from the general one in the case where a single region is considered—in which case instruments reduce to POVMs and process matrices reduce to density matrices [20].

We are in particular interested in the situation where two consecutive measurements are performed on a single quantum system. Gleason-type derivations of quantum probabilities do not tell us how to assign joint probabilities to two such events: one must introduce an additional ingredient—the state update rule. If the statistics for the first measurement are described by a density matrix ρ, and the first measurement is described by a CP map $\mathcal{M}$, one calculates the probabilities for the second measurement, given the outcome of the first is known, by applying the Born rule to the updated state [51]

$$\rho \rightsquigarrow \widetilde{\rho}_{\mathcal{M}} = \frac{\mathcal{M}(\rho)}{\mathrm{Tr}\,\mathcal{M}(\rho)} = \frac{\sum_{j=1}^{m} K_j \rho K_j^{\dagger}}{\mathrm{Tr}\left(\sum_{j=1}^{m} K_j^{\dagger} K_j \rho\right)}. \tag{17}$$

(Note that the update rule does not depend on the particular decomposition of $\mathcal{M}$ into Kraus operators $\{K_j\}_{j=1}^{m}$.) In an operational perspective, rule (17) is seen as a quantum analogue of classical knowledge update. Within the quantum process framework, this is more than an analogy: the update rule is *derived* from the joint probability assignment.

To make the argument rigorous, we should remark again that the quantum frame function *is not* a normalised probability measure over the entire space of potential events. Formally, the frame function defines a parametrised probability for observing a CP map $\mathcal{M}^X$ given an instrument $\mathfrak{I}^X$ in regions $X = A, ...$:

$$P(\mathcal{M}^A, ...|\mathfrak{I}^A, ...) = f(\mathcal{M}^A, ...) \quad \text{if } \mathcal{M}^A \in \mathfrak{I}^A, ...$$
$$= 0 \quad \text{otherwise.} \tag{18}$$

(This defines a conditional probability if a marginal $P(\mathfrak{I}^A, ...)$ is assigned.) Even though the inclusion of instruments is necessary to define (18) as a classical probability, we will omit them in the following for notational convenience.

Expression (18) defines an ordinary, classical probability measure, which lets us use all the machinery of classical probability theory. In particular, the conditional probability to observe $\mathcal{M}^B$ in region B, given that $\mathcal{M}^A$ is observed in region A, can be calculated from the joint probability distribution:

$$\begin{aligned} P(\mathcal{M}^B|\mathcal{M}^A) &= \frac{P(\mathcal{M}^B, \mathcal{M}^A)}{P(\mathcal{M}^A)} \\ &= \frac{\mathrm{Tr}[(M^A \otimes M^B) \cdot W]}{\sum_{M^B \in \mathfrak{I}^B} \mathrm{Tr}[(M^A \otimes M^B) \cdot W]} \\ &= \mathrm{Tr}(M^B\, \widetilde{W}_{M^A}), \end{aligned} \tag{19}$$

where we introduced the *conditional process matrix* [44]

$$\widetilde{W}_{M^A} := \frac{\mathrm{Tr}_{A_I A_O}[(M^A \otimes \mathbf{1}^B) \cdot W]}{\mathrm{Tr}\left[\left(M^A \otimes \sum_{M^B \in \mathfrak{I}^B} M^B\right) \cdot W\right]}. \tag{20}$$

Relevant to the ordinary state update rule is the case where A precedes temporally B, and the evolution between the two events is trivial. This scenario is described by the process matrix (see, e.g., [22])

$$W = \rho^{A_I} \otimes [[\mathbf{1}]]^{A_O B_I} \otimes \mathbf{1}^{B_O}, \tag{21}$$

$$[[\mathbf{1}]]^{A_O B_I} := \sum_{jl} |j\rangle \langle l|^{A_O} \otimes |j\rangle \langle l|^{B_I}, \tag{22}$$

where ρ is the density matrix describing the input state of region A. A straightforward calculation shows that, in this case, the conditional process matrix reduces to

$$\widetilde{W}_{M^A} = \left[\frac{\mathcal{M}^A(\rho)}{\mathrm{Tr}\,\mathcal{M}^A(\rho)} \right]^{B_I} \otimes \mathbb{1}^{B_O} \equiv \widetilde{\rho}_{\mathcal{M}^A} \otimes \mathbb{1}^{B_O}, \qquad (23)$$

which is the process matrix description of region *B* receiving a state described by the density matrix $\widetilde{\rho}_{\mathcal{M}^A}$.

Conditioning versus updating

Some clarification at this point might be helpful regarding the distinction between *probabilistic conditioning* and *knowledge update* (see, e.g., [52] for a more detailed discussion). In classical probability theory, the rules for probabilistic conditioning simply follow from the axioms of the theory. These axioms can be assumed, as in Kolmogorov's approach, or derived from requirements on how one should consistently assign degrees of belief, for example through Dutch book arguments [53] or theorems like that of Savage [54]. A consequence of probabilistic conditioning is Bayes' theorem for inverting conditional probabilities:

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}.$$

Importantly, this rule does not involve updating knowledge given *new* information: all information is contained in the joint probability $P(A, B)$, which is unchanged by the application of the theorem.

Knowledge update, on the other hand, refers to the process of updating one's belief following the acquisition of new data. This process is not encoded in the axioms of probability theory and requires extra assumptions. For example, if one assumes that all data values that were not observed can be discarded, one arrives at Bayes rule:

$$P_{\mathrm{new}}(A) = P_{\mathrm{old}}(A|B = b),$$

where *b* is the observed value of the variable *B*. Implicit in this rule is the counterfactual assumption that values that are not observed are known to be false. Thus, applying Bayes theorem, one arrives at the standard form for Bayes updating:

$$P_{\mathrm{new}}(A) = P_{\mathrm{old}}(A)\frac{P_{\mathrm{old}}(B = b|A)}{P_{\mathrm{old}}(B = b)}.$$

In the quantum case, such counterfactual assumptions are known to be problematic [55][6]. In our approach, no such assumption is necessary, because the primitive object is the joint probability. From this perspective, the 'state update' rule is not an update at all, but rather an application of probabilistic conditioning. This insight distinguishes this approach from other attempts to leverage Bayesian arguments to justify an informational interpretation of quantum theory [21, 23].

Discussion

In this work we have shown that it is possible to use a Gleason-type approach to derive a quantum probability rule that subsumes both the Born rule and the state update rule. By using the structure of local quantum operations and a reasonable non-contextuality assumption we have derived both the new rule and the appropriate object to represent the arbitrary background structure, or process. The central feature of the probability rule is linearity. In contrast to previous derivations, where linearity was assumed [17, 20], here we have shown that it can be derived from the assumption of non-contextuality alone[7].

Our demonstration that the state update, or 'collapse' rule can be regarded as non-fundamental offers a new perspective on a variety of foundational questions. In particular, informational interpretations of wavefunction collapse can now be given a rigorous foundation: state update can be viewed as a case of classical probabilistic conditioning. The work here also presents the opportunity to extend no-go theorems for non-contextual hidden variable models to scenarios involving more general causal structures [56].

Finally, a key advantage of the approach presented here is that it does not presuppose any *a priori* distinction between space-like and time-like separated events. Therefore, it avoids conceptual difficulties associated with the non-covariant nature of the state update rule. It is thus a promising direction for research aimed at developing a fully relativistic version of the formalism that encodes space-time symmetries.

---

[6] That is not to say that one can not apply counterfactual reasoning successfully in the case of single quantum contexts. Indeed, it is only when one extends the requirement to a joint probability over all contexts that such counterfactual reasoning becomes problematic.

[7] Obtaining linearity of the probability rule via an extension of Gleason's theorem was suggested, but not proved, in the supplementary methods of [20].

Acknowledgments

Appendix. Inner product for linear maps

Here we construct the inner product on the space of linear maps $L^A = \{\mathcal{M}: \mathcal{L}(\mathcal{H}^{A_I}) \to \mathcal{L}(\mathcal{H}^{A_O})\}$ and derive its CJ representation. Recall that, given an inner product $\langle\psi|\phi\rangle$ on a Hilbert space $\mathcal{H}$ and an arbitrary basis that is orthonormal with respect to this product, $\langle e_j|e_k\rangle = \delta_{jk}$, one defines the Hilbert–Schmidt scalar product for operators $\sigma, \rho \in \mathcal{L}(\mathcal{H})$ as

$$(\rho, \sigma)_{\mathrm{HS}} := \sum_k \langle\rho(e_k)|\sigma(e_k)\rangle = \mathrm{Tr}(\rho^\dagger\sigma), \tag{24}$$

where we momentarily abandon the Dirac notation and represent explicitly the action of an operator on a vector as $v \in \mathcal{H} \rightsquigarrow \rho(v) \in \mathcal{H}$. Note that the Hilbert–Schmidt inner product does not depend on which basis is used in its definition, as long as it is orthonormal with respect to the underlying Hilbert space inner product.

We move a step further and, based on the Hilbert–Schmidt inner product, define an inner product for the space L^A of linear maps. For this purpose, we select a basis of Hermitian matrices for the input space that is orthonormal with respect to the Hilbert–Schmidt product (called Hilbert–Schmidt basis):

$$\tau_\mu \in \mathcal{L}(\mathcal{H}^{A_I}),$$
$$\tau_\mu = \tau_\mu^\dagger,$$
$$\mathrm{Tr}(\tau_\mu \tau_\nu) = \delta_{\mu\nu}.$$

The inner product between any two linear maps $\mathcal{M}$, $\mathcal{N}$ is then defined in analogy to equation (24) and coincides with the inner product introduced in the main text:

$$\begin{aligned}(\mathcal{M}, \mathcal{N})_S &:= \sum_\mu (\mathcal{M}(\tau_\mu), \mathcal{N}(\tau_\mu))_{\mathrm{HS}} \\ &= \sum_\mu \mathrm{Tr}[\mathcal{M}(\tau_\mu)^\dagger \mathcal{N}(\tau_\mu)],\end{aligned} \tag{25}$$

where the subscript S stands for 'superoperator'. Note that, just as for equation (24), the RHS of equation (25) can be rewritten as a superoperator trace and is thus independent of the choice of basis.

Next, we want to relate the superoperator inner product to the CJ representation. Reintroducing the Dirac notation, the CJ inner product between operators is defined as

$$\mathcal{M} \to M := \left[\sum_{jk} |j\rangle\langle k|^{A_I} \otimes \mathcal{M}(|j\rangle\langle k|)^{A_O}\right]^T, \tag{26}$$

$$(\mathcal{M}, \mathcal{N})_{\mathrm{CJ}} := \mathrm{Tr}(M^\dagger N). \tag{27}$$

Note that the inner product keeps the same form if definition (26) is replaced by its transpose. We can thus re-write it as

$$\begin{aligned}(\mathcal{M}&, \mathcal{N})_{\mathrm{CJ}} \\ &= \sum_{jkmn} \mathrm{Tr}[|j\rangle\langle k|^{A_I} \otimes \mathcal{M}(|j\rangle\langle k|)^{A_O}]^\dagger |m\rangle\langle n|^{A_I} \otimes \mathcal{N}(|m\rangle\langle n|)^{A_O} \\ &= \sum_{jkmn} \langle j|m\rangle\langle n|k\rangle \mathrm{Tr}[\mathcal{M}(|j\rangle\langle k|)^\dagger \mathcal{N}(|m\rangle\langle n|)] \\ &= \sum_{mk} \mathrm{Tr}[\mathcal{M}(|m\rangle\langle k|)^\dagger \mathcal{N}(|m\rangle\langle k|)].\end{aligned} \tag{28}$$

To see how this relates to the superoperator inner product, we need to recall two useful facts.

Lemma 3. *Given a Hilbert space* $\mathcal{H}$, *the* swap operator $S\colon \mathcal{H} \otimes \mathcal{H} \to \mathcal{H} \otimes \mathcal{H}$, *defined by its action* $S|\psi\rangle|\phi\rangle = |\phi\rangle|\psi\rangle$, *can be written as*

$$S = \sum_{\mu} \tau_{\mu} \otimes \tau_{\mu} \tag{29}$$

for an arbitrary Hilbert–Schmidt basis $\{\tau_{\mu}\} \subset \mathcal{L}(\mathcal{H})$.

Proof. Viewed as an operator, S can be decomposed with respect to a basis $\{|j\rangle\}$ of the Hilbert space $\mathcal{H}$ as $S = \sum_{km}|k\rangle\langle m| \otimes |m\rangle\langle k|$. On the other hand, viewed as a vector on the linear space of operators $\mathcal{L}(\mathcal{H} \otimes \mathcal{H})$, S can be decomposed with respect to the Hilbert–Schmidt basis as

$$S = \sum_{\mu\nu} \tau_{\mu} \otimes \tau_{\nu} \operatorname{Tr}\left[(\tau_{\mu} \otimes \tau_{\nu}) \cdot S\right]. \tag{30}$$

The components in the above representation are given by

$$
\begin{aligned}
\operatorname{Tr}[(\tau_{\mu} \otimes \tau_{\nu}) \cdot S] &= \sum_{km} \operatorname{Tr}[(\tau_{\mu} \otimes \tau_{\nu}) \cdot (|k\rangle\langle m| \otimes |m\rangle\langle k|)] \\
&= \sum_{km} \langle m|\tau_{\mu}|k\rangle \langle k|\tau_{\nu}|m\rangle \\
&= \operatorname{Tr}(\tau_{\mu}\tau_{\nu}) = \delta_{\mu\nu}.
\end{aligned}
$$

Plugging this into the decomposition (30), we obtain equation (29). $\qquad\square$

This lemma can be used to prove the completeness relation

$$\sum_{\mu} \langle m|\tau_{\mu}|k\rangle^{*} \langle n|\tau_{\mu}|r\rangle = \delta_{mn}\delta_{kr}. \tag{31}$$

Indeed, using $\tau_{\mu} = \tau_{\mu}^{\dagger}$, we have

$$
\begin{aligned}
&\sum_{\mu} \langle m|\tau_{\mu}|k\rangle^{*} \langle n|\tau_{\mu}|r\rangle \\
&= \sum_{\mu} \langle k|\langle n|\tau_{\mu} \otimes \tau_{\mu}|m\rangle|r\rangle \\
&= \langle k|\langle n|S|m\rangle|r\rangle = \delta_{mn}\delta_{kr}.
\end{aligned}
$$

We can now re-write the superoperator inner product:

$$
\begin{aligned}
(\mathcal{M}, \mathcal{N})_{\mathrm{S}} &= \sum_{\mu} \operatorname{Tr}[\mathcal{M}(\tau_{\mu})^{\dagger}\mathcal{N}(\tau_{\mu})] \\
&= \sum_{mknr} \sum_{\mu} \langle m|\tau_{\mu}|k\rangle^{*} \langle n|\tau_{\mu}|r\rangle \operatorname{Tr}\mathcal{M}(|m\rangle\langle k|)^{\dagger}\mathcal{N}(|n\rangle\langle r|) \\
&= \sum_{km} \operatorname{Tr}[\mathcal{M}(|m\rangle\langle k|)^{\dagger}\mathcal{N}(|m\rangle\langle k|)].
\end{aligned}
$$

Comparing this with equation (28), we conclude that $(\mathcal{M}, \mathcal{N})_{\mathrm{CJ}} = (\mathcal{M}, \mathcal{N})_{\mathrm{S}}$.

ORCID iDs

Sally Shrapnel ⓘ https://orcid.org/0000-0001-8407-7176
Fabio Costa ⓘ https://orcid.org/0000-0002-6547-6005

References

[1] Ghirardi G C, Rimini A and Weber T 1986 Unified dynamics for microscopic and macroscopic systems *Phys. Rev.* D **34** 470
[2] Tumulka R 2006 On spontaneous wave function collapse and quantum field theory *Proc. R. Soc.* A **462** 1897–908
[3] Bassi A, Lochan K, Satin S, Singh T P and Ulbricht H 2013 Models of wave-function collapse, underlying theories, and experimental tests *Rev. Mod. Phys.* **85** 471–527
[4] Gisin N 2017 Collapse. What else? arXiv:1701.08300
[5] Ozawa M 1997 Quantum state reduction and the quantum Bayes principle *Quantum Communication, Computing, and Measurement* (Berlin: Springer) pp 233–41
[6] Fuchs C A 2002 Quantum mechanics as quantum information (and only a little more) arXiv:quant-ph/0205039
[7] Caves C M, Fuchs C A and Schack R 2002 Quantum probabilities as Bayesian probabilities *Phys. Rev.* A **65** 022305
[8] Wiseman H M and Milburn G J 2009 *Quantum Measurement and Control* (Cambridge: Cambridge University Press)
[9] Timpson C G 2013 *Quantum Information Theory and the Foundations of Quantum Mechanics* (Oxford: Oxford University Press)
[10] Hellmann F, Kamiński W and Kostecki R P 2016 Quantum collapse rules from the maximum relative entropy principle *New J. Phys.* **18** 013022

[11] Mermin N D 2017 Why QBism is not the Copenhagen interpretation and what John Bell might have thought of it *Quantum (Un) speakables II* (Berlin: Springer)

[12] Brukner C 2017 On the Quantum measurement problem *Quantum (Un)speakables II* (Berlin: Springer)

[13] Leifer M S 2006 Quantum dynamics as an analog of conditional probability *Phys. Rev.* A **74** 042310

[14] Gutoski G and Watrous J 2006 Toward a general theory of quantum games *Proc. 39th ACM STOC (San Diego, CA, 11–13 June, 2007)* pp 565–74

[15] Chiribella G, D'Ariano G M and Perinotti P 2008 Quantum circuit architecture *Phys. Rev. Lett.* **101** 060401

[16] Chiribella G, D'Ariano G M and Perinotti P 2008 Memory effects in quantum channel discrimination *Phys. Rev. Lett.* **101** 180501

[17] Chiribella G, D'Ariano G M and Perinotti P 2009 Theoretical framework for quantum networks *Phys. Rev.* A **80** 022339

[18] Bisio A, Chiribella G, D'Ariano G and Perinotti P 2011 Quantum networks: general theory and applications *Acta Phys. Slovaca. Rev. Tutorials* **61** 273–390

[19] Bisio A, D'Ariano G M, Perinotti P and Sedlák M 2014 Optimal processing of reversible quantum channels *Phys. Lett.* A **378** 1797–808

[20] Oreshkov O, Costa F and Brukner Č 2012 Quantum correlations with no causal order *Nat. Commun.* **3** 1092

[21] Leifer M S and Spekkens R W 2013 Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference *Phys. Rev.* A **88** 052130

[22] Costa F and Shrapnel S 2016 Quantum causal modelling *New J. Phys.* **18** 063032

[23] Allen J-M A, Barrett J, Horsman D C, Lee C M and Spekkens R W 2017 Quantum common causes and quantum causal models *Phys. Rev.* X **7** 031021

[24] Modi K 2012 Operational approach to open dynamics and quantifying initial correlations *Sci. Rep.* **2** 581

[25] Ringbauer M, Wood C J, Modi K, Gilchrist A, White A G and Fedrizzi A 2015 Characterizing quantum dynamics with initial system-environment correlations *Phys. Rev. Lett.* **114** 090402

[26] Pollock F A, Rodríguez-Rosario C, Frauenheim T, Paternostro M and Modi K 2018 Non-Markovian quantum processes: complete framework and efficient characterization *Phys. Rev.* A **97** 012127

[27] Milz S, Pollock F A and Modi K 2016 Reconstructing open quantum system dynamics with limited control arXiv:1610.02152

[28] Gleason A 1957 Measures on the closed subspaces of a Hilbert space *J. Math. Mech.* **6** 885–93

[29] Busch P 2003 Quantum states and generalized observables: a simple proof of Gleason's theorem *Phys. Rev. Lett.* **91** 120403

[30] Caves C M, Fuchs C A, Manne K K and Renes J M 2004 Gleason-type derivations of the quantum probability rule for generalized measurements *Found. Phys.* **34** 193–209

[31] Barnett S M, Cresser J D, Jeffers J and Pegg D T 2014 Quantum probability rule: a generalization of the theorems of Gleason and Busch *New J. Phys.* **16** 043025

[32] Hall M J 2016 Comment on 'Gleason-type theorem for projective measurements, including qubits' by F. De Zela arXiv:1611.00613

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

[34] Mazurek M D, Pusey M F, Kunjwal R, Resch K J and Spekkens R W 2016 An experimental test of noncontextuality without unphysical idealizations *Nat. Commun.* **7** 11780

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

[36] Nielsen M and Chuang I 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)

[37] Hellwig K E and Kraus K 1969 Pure operations and measurements *Commun. Math. Phys.* **11** 214–20

[38] Hellwig K E and Kraus K 1970 Operations and measurements. II *Commun. Math. Phys.* **16** 142–7

[39] Davies E and Lewis J 1970 An operational approach to quantum probability *Commun. Math. Phys.* **17** 239–60

[40] Smith R T and Minton R B 1955 *Calculus* (New York: McGraw-Hill)

[41] Brukner Č 2014 Quantum causality *Nat. Phys.* **10** 259–63

[42] Araújo M, Branciard C, Costa F, Feix A, Giarmatzi C and Brukner Č 2015 Witnessing causal nonseparability *New J. Phys.* **17** 102001

[43] Feix A, Araújo M and Brukner Č 2015 Quantum superposition of the order of parties as a communication resource *Phys. Rev.* A **92** 052326

[44] Oreshkov O and Giarmatzi C 2016 Causal and causally separable processes *New J. Phys.* **18** 093020

[45] Branciard C, Araújo M, Feix A, Costa F, Costa F and Brukner Č 2016 The simplest causal inequalities and their violation *New J. Phys.* **18** 013008

[46] Baumann V and Brukner Č 2016 Appearance of causality in process matrices when performing fixed-basis measurements for two parties *Phys. Rev.* A **93** 062324

[47] Oreshkov O and Cerf N J 2016 Operational quantum theory without predefined time *New J. Phys.* **18** 073037

[48] Abbott A A, Giarmatzi C, Costa F and Branciard C 2016 Multipartite causal correlations: polytopes and inequalities *Phys. Rev.* A **94** 032131

[49] Choi M-D 1975 Completely positive linear maps on complex matrices *Linear Algebra Appl.* **10** 285–90

[50] Jamiołkowski A 1972 Linear transformations which preserve trace and positive semidefiniteness of operators *Rep. Math. Phys.* **3** 275–8

[51] Kraus K 1971 General state changes in quantum theory *Ann. Phys.* **64** 311–35

[52] Caticha A and Giffin A 2006 Updating probabilities *AIP Conf. Proc.* **872** 31–42

[53] De Finetti B 1990 *Theory of Probability* (New York: Wiley)

[54] Savage L J 1954 *The Foundations of Statistics* (*Wiley Publications in Statistics*) (New York: Wiley)

[55] Żukowski M and Brukner Č 2014 Quantum non-locality—it ain't necessarily so... *J. Phys. A: Math. Theor.* **47** 424009

[56] Shrapnel S and Costa F 2017 Causation does not explain contextuality arXiv:1708.00137