The logic and structure of quantum stochastic processes (and causal modelling)

Simon Milz,[1,*] Fattah Sakuldee,[2] Felix A. Pollock,[1] and Kavan Modi[1]

[1] *School of Physics and Astronomy, Monash University, Clayton, Victoria 3800, Australia*
[2] *MU-NECTEC Collaborative Research Unit on Quantum Information, Department of Physics, Faculty of Science, Mahidol University, Bangkok 10400, Thailand.*

(Dated: March 31, 2019)

In classical physics, the Kolmogorov extension theorem lays the foundation for the theory of stochastic processes. It has been known for a long time that, in its original form, this theorem does not hold in quantum mechanics. More generally, it does not hold in any theory of stochastic processes – classical, quantum or beyond – that does not just describe passive observations, but allows for active interventions. Such processes form the basis of the study of causal modelling across the sciences, including in the quantum domain. To date, the associated frameworks lack a firm theoretical underpinning. We prove a generalized extension theorem that applies to all theories of stochastic processes, putting them on equally firm mathematical ground as their classical counterpart.

Along the way we show that quantum causal modelling and quantum stochastic processes are equivalent. This provides the correct framework for the description of experiments involving continuous control, which play a crucial role in the development of quantum technologies. Furthermore, we show that the original extension theorem follows from the generalized one in the correct limit, and elucidate how a comprehensive understanding of general stochastic processes allows one to unambiguously define the distinction between those that are classical and those that are quantum.

I. INTRODUCTION

Stochastic processes are ubiquitous in nature. Their theory is used, among other applications, to model the stock market, predict the weather, describe transport processes in cells and understand the random motion of particles suspended in a fluid [1, 2]. Intuitively, when we speak of stochastic processes, we often mean joint probability distributions of random variables at a finite set of times: the probability for a stock to have prices P_1, P_2 and P_3 on three subsequent days, or the probability to find a particle undergoing Brownian motion in regions R_1 and R_2 when measuring its position at times t_1 and t_2.

This finite description of stochastic processes is motivated by both experimental and mathematical considerations. On the experimental side, temporal resolution is generally limited and digital instruments always record a finite amount of data. Hence, the only accessible information we are left with is encoded in probability distributions with a finite number of arguments. On the mathematical side, it is much less cumbersome to model stochastic processes on discrete times – for example, by defining transition probabilities $\mathbb{P}(Y|X)$ between random variables at a fixed set of different times – than modelling probability densities on the space of all possible 'trajectories' of random variables.

These motivations notwithstanding, the fundamental laws of physics are continuous in nature and one always implicitly assumes that there is an *underlying* process that leads to the experimentally accessible finite distributions. Put differently, one assumes that there exists an infinite joint probability distribution that has *all* the finite ones as marginals. For classical stochastic processes, these two points of view, the finite and the infinite one, are reconciled by the Kolmogorov extension theorem (**KET**), which lays bare the minimal requirements for the existence of an underlying process, given a family of measurement statistics for finite sets of times [3–6]. It bridges the gap between experimental reality and mathematically rigorous theoretical underpinnings and, as such, enables the definition of stochastic processes as the limit of finite – and hence observable – objects. Additionally, the KET enables the modelling of continuous processes based on finite probability distributions. As a consequence, in the classical setting, stochastic processes over a continuous set of times, and families of finite probability distributions are two sides of the same coin.

The validity of the KET hinges crucially on the fact that the statistics of observations at a time t do not depend on the kind of measurements that were made at any time $t' < t$. Put differently, just like the Leggett-Garg inequalities for temporal correlations [7–9], the KET is based on the assumption of macroscopic realism and noninvasive measurements. For example, in a classical stochastic process, measuring the position of a particle undergoing Brownian motion, merely reveals information, but does not actively change the state of the particle.

On the other hand, the assumptions of non-invasiveness or macroscopic realism are not fulfilled in many experimental scenarios, leading to a breakdown of the KET, at the cost of a clear connection between an underlying process and its finite time manifestations.

* simon.milz@monash.edu

This is the case whenever an experimenter chooses to actively interfere with a process to uncover its causal structure or to investigate the reaction to different inputs. For example, instead of just observing the progress of a disease, a pharmacologist tries to find out how the course of a disease changes with the administration of certain drugs. More generally, agent based modelling investigates how systems behave when they can not only be monitored, but actively influenced [10]. Experimental situations where interventions are actively used to uncover causal relations fall within the field of *causal modelling* [11].

Interventions appear naturally in quantum mechanics, where generic measurements necessarily perturb a system's state; in fact, a complete description of quantum processes without interventions is not possible [12]. As in the classical case, interventions can also be used to actively probe the causal structure of a quantum process, and the description of quantum processes with interventions has been recently used to develop the field of *quantum causal modelling* [13–15]. Importantly, as in the case of classical processes with interventions, the invasiveness of measurements means that the KET does not hold for quantum processes [16]. This is analogous to the violation of Leggett-Garg inequalities [7, 9] in quantum mechanics.

The fundamental lack of an extension theorem in quantum mechanics (or any other theory with interventions) would be problematic for several reasons: Firstly, it would suggest a lack of consistency between descriptions of a process for different sets of times; for example, the description of a process for three times t_1, t_2, and t_3 would *not* already include the description of the process for the two times t_1 and t_3 only. In other words, we would need seven different independent descriptors for each of the seven subsets of times to describe all possible events! This lack of consistency would render the study of (quantum) causal models in multi-step experiments impossible; if local interventions lead to a completely different process, it is not meaningful to try to deduce causal relations by means of active manipulations of the system at hand.

Furthermore, the present situation (with no extension theorem) implies an incompatibility between existing frameworks to describe processes with interventions (both classical and quantum) and the classical theory of stochastic processes, even though they should converge to the latter in the correct limit. This then suggests that the mere act of interacting with a system over time introduces a fundamental divide between the continuity of physical laws and the finite statistics that can be accessed in reality, thus begging the question: What do we generally mean by a (quantum) 'stochastic process', and how can we reconcile causal modelling frameworks with the idea of an underlying process?

In this Paper, we answer these questions by generalizing the KET to the framework of (quantum) causal modelling, thus closing the apparent divide between the finite and the continuous point of view. To this end, in Sec. II we reiterate the relation between classical stochastic processes and classical causal modelling and show the breakdown of the KET when we allow for active interventions in Sec. III. We analyze the quantum case in Sec. IV and find that stochastic processes can only be defined properly by taking interventions into account. Consequently, the framework of quantum stochastic processes is equivalent to quantum causal modelling. In Sec. IV B, we prove our main result, that the KET can be generalized to quantum stochastic processes, and this *generalized extension theorem* (**GET**) reduces to the classical one in the correct limit. The breakdown of the KET is a breakdown of formalism only, not a fundamental property of quantum processes. Our generalized extension theorem provides an overarching theorem that puts all processes with interventions and, in particular, quantum processes on an equally sound footing as their classical counterpart.

We discuss the equivalence of quantum stochastic processes and quantum causal modelling in Sec. IV C. As a direct application, in Sec. IV D, we use the GET to provide a distinction between general, *i.e.*, non-Markovian, classical and quantum processes, as has been recently introduced in [17] for the restricted case of processes without memory. While we phrase our results predominantly in the language of causal modelling, they apply to a wide range of current theories of quantum processes and beyond. The relation of our results to other frameworks, in particular to the work of Accardi, Frigerio and Lewis [18], is discussed in Sec. V. We conclude the Paper in Sec. VI.

II. CLASSICAL STOCHASTIC PROCESSES AND CAUSAL MODELLING

A. Classical stochastic processes

A classical stochastic process can be described by joint probability distributions $\mathbb{P}_{\Lambda_k}(i_k, \dots, i_1) := \mathbb{P}_{\Lambda_k}(\mathbf{i}_{\Lambda_k})$ of random variables that take values $\{i_\alpha\}$ at time t_α [5], where Λ_k is a collection of times with cardinality $|\Lambda_k| = k$. For instance, for a k step process, the set of times could be $\Lambda_k = \{t_k, \dots, t_1\}$. We will employ the convention that subscripts signify the time as well as the particular value of the respective random variable. For example, i_α signifies a value of the random variable at time t_α.

The distribution $\mathbb{P}_{\Lambda_k}(\mathbf{i}_k)$ could express the probability for a particular length-k sequence of heads and tails when flipping a coin, or the probability to find a particle undergoing Brownian motion at positions $\mathbf{i}_k$ when measuring it at times Λ_k. Importantly, this description of a stochastic process is sufficient to describe the behaviour on any subset of the times considered; for instance, the distribution over all but the jth time is found by marginalizing the larger distribution: $\mathbb{P}_{\Lambda_k\setminus\{t_j\}}(i_k, \dots, i_{j+1}, i_{j-1}, \dots, i_1) = \sum_{i_j} \mathbb{P}_{\Lambda_k}(\mathbf{i}_k)$. This property implicitly assumes that there is only one instrument that is used to interrogate the system of interest, and this interrogation does not influence its state. Neither of these assumption are fulfilled in more general processes, such as the ones employed in causal modelling.

B. Classical causal modelling

Observing the statistics for measurement outcomes reveals correlations between events, but no information about causal relations. For instance, correlations of two events A and B could stem from A influencing B, B influencing A, or both A and B being influenced by an earlier event C [11, 13, 15] (see Fig. 1). Reiterating an example from Ref. [15], events A and B could be the occurrence of sunburns and the sales of ice cream, respectively. While these two variables are highly correlated, this correlation alone would not fix a causal relation between them. Inferring the causal structure of a process is the aim of *causal modelling*. Here, active interventions are used to uncover how different events can influence each other. In the example above, one could suspend the sale of ice cream to see how it affects the occurrence of sunburns, and would find out that ice cream sales have no direct effect on sunburns (and vice versa, as the correlations of ice cream sales and sunburns stem from a common cause, the weather, and not from any direct causal relation).

Mathematically, causal modelling for k events $A_k, \dots, A_1$ necessitates the collection of all joint probability distributions $\mathbb{P}_{\Gamma_k}(i_k, \dots, i_1|j_k, \dots, j_1) := \mathbb{P}_{\Gamma_k}(\mathbf{i}_{\Gamma_k}|\boldsymbol{\mathcal{J}}_{\Gamma_k})$ to measure the outcomes $i_k, \dots, i_1$ given that the *interventions* $j_k, \dots, j_1$ were performed. Here, Γ_k is a set of labels for events; *a priori*, there is no particular order imposed on the elements of Γ_k, and we use a different letter for the set of event labels to distinguish it from the set of times Λ_k used above. For example, Γ_k could contain labels for different laboratories where experiments are performed. $\boldsymbol{\mathcal{J}}_{\Gamma_k}$ are the *instruments* that were used at each of the events; these can be seen as rules for how to intervene upon seeing a particular outcome (we will formalize the notion of an instrument in Sec. IV). For example, when investigating Brownian motion, an instrument could be a deterministic replacement rule: upon finding the particle at i_a replace it by a particle at i'_a. It could also be probabilistic: upon finding the particle at i_a, with probability $p_{i'_a}$ replace it by a particle at i'_a.

One possible instrument is the trivial idle instrument $\mathcal{J}_a = \mathrm{id}_a$, the instrument that only measures the particle but doesn't change it. For classical stochastic processes, the corresponding joint distribution over outcomes can be thought of as the instrument-independent underlying distribution of the random variables describing the process:

$$\mathbb{P}_{\Gamma_k}(\mathbf{i}_{\Gamma_k}|\mathbf{id}_{\Gamma_k}) = \mathbb{P}_{\Gamma_k}(\mathbf{i}_{\Gamma_k})\,, \tag{1}$$

where $\mathbf{id}_{\Gamma_k}$ denotes the idle instrument at each of the events in Γ_k. If we chose Λ_k to be a set of times Λ_k, the right-hand side of Eq. (1) has the form of a k-step stochastic process. This directly leads to the following (well-known) proposition:

Proposition 1. *Classical causal modelling contains classical stochastic processes as a special case.*

As mentioned, this statement follows by choosing the set of events Γ_k to be a set Λ_k of ordered times and the instruments to be the idle instrument. We emphasize that causal modelling does not impose a temporal ordering *per se*, but deduces the ordering of events from the obtained correlation functions (finding this order, or, more precisely, the underlying *directed acyclic graph* (DAG) that defines the causal relations of the events, is the original aim of causal modelling [11, 13]). As neither the proof of the KET, nor the proof of the GET makes use of the notion of *a priori* temporal ordering (see Sec. IV B for a discussion), in what follows, we will drop the distinction between sets of labels Γ_k and sets of times Λ_k. We now show that the introduction of interventions, that is crucial for the deduction of causal relations, leads to a breakdown of fundamental properties that are satisfied by classical stochastic processes.

III. THE KOLMOGOROV EXTENSION THEOREM AND INTERVENTIONS

A. The KET

The KET is concerned with the question of which properties a family of finite joint probability distributions have to satisfy in order for an underlying process to exist. As such, it defines the classical notion of a stochastic process. In what follows, we will distinguish between stochastic processes on a finite number of times – which are characterized by joint probability distributions with finitely many arguments – and the *underlying* stochastic process that leads to *all* of these finite distributions.

As already mentioned, a classical stochastic process is described by a family of joint probability distributions $\mathbb{P}_{\Lambda_k}(\mathbf{i}_{\Lambda_k})$ for different finite sets of times Λ_k. An underlying process on a set Λ (finite, countably or uncountably infinite) is a joint probability distribution $\mathbb{P}_\Lambda$, that has all finite ones as marginals. In detail, we have $\mathbb{P}_{\Lambda_k}(\mathbf{i}_{\Lambda_k}) = \sum_{\Lambda\setminus\Lambda_k} \mathbb{P}_\Lambda(\mathbf{i}_\Lambda) := \mathbb{P}_\Lambda^{|\Lambda_k}(\mathbf{i}_{\Lambda_k})$ for all $\Lambda_k \subseteq \Lambda$, where $\mathbf{i}_{\Lambda_k}$ is the subset of $\mathbf{i}_\Lambda$ corresponding to the times Λ_k, $\sum_{\Lambda\setminus\Lambda_k}$ denotes a sum over realizations of the random variables at all times that are part of $\Lambda \setminus \Lambda_k$ (*i.e.*, all the times that lie in Λ but not in Λ_k), and $\mathbb{P}_\Lambda^{|\Lambda_k}$ denotes the restriction of $\mathbb{P}_\Lambda$ to the times Λ_k. In

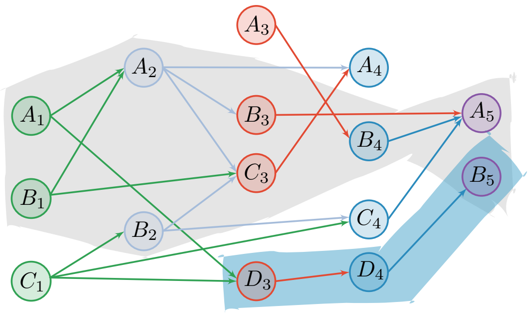

Figure 1. *(Quantum) Causal network.* Performing different interventions allows for the causal relations between events to be probed. For example, in the figure the event B_1 directly influences the events C_3 and A_2, while A_3 influences only B_4. Depending on the degrees of freedom that can be accessed by the experimenter, these causal relations can or cannot be detected. For example, the influence of A_3 on B_4 could not be discovered if only the degrees of freedom in the gray area were experimentally accessible. Independent of the accessible degrees of freedom, the GET holds for *any* process. On the other hand, the statistics of events do in general not satisfy the requirements of the KET. For example, the events D_3, D_4, B_5 could be successive (*e.g.*, at times t_3, t_4 and t_5) spin measurements in z-, x- and z-direction, respectively. Summing over the results of the spin measurement in x-direction at t_4 would not yield the correct probability distribution for two measurements in z-direction at t_3 and t_5 only (see also Sec. IV A).

the case where the set Λ is infinite, the marginalization procedure can correspond to an integral over the times in $\Lambda \setminus \Lambda_k$ (though, to avoid introducing too much notation, we will still use $\sum_{\Lambda\setminus\Lambda_k}$ to represent it). For example, if the process we are interested in is the Brownian motion of a particle, $\mathbb{P}_\Lambda$ would be the probability density of all possible trajectories that the particle could take in the time interval Λ, and all finite distributions could in principle be obtained from $\mathbb{P}_\Lambda$.

If the finite joint probability distributions stem from an underlying process, it is easy to see that probability distributions for any two finite subsets of times $\Lambda_k \subseteq \Lambda_\ell \subseteq \Lambda$ fulfill a *consistency condition* (or *compatibility condition*) amongst each other, *i.e.*, $\mathbb{P}_{\Lambda_k}$ is a marginal of $\mathbb{P}_{\Lambda_\ell}$. Expressed in the notation introduced above, we have $\mathbb{P}_{\Lambda_k} = \mathbb{P}_{\Lambda_\ell}^{|\Lambda_k}$ for all $\Lambda_k \subseteq \Lambda_\ell \subseteq \Lambda$. Intuitively, this means that $\mathbb{P}_{\Lambda_\ell}$, the descriptor of the stochastic process on the times Λ_ℓ, contains all information about subprocesses on fewer times.

While an underlying process leads to a family of compatible finite probability distributions, the KET shows that the converse is also true. Any family of consistent probability distributions implies the existence of an underlying process. Specifically, the Kolmogorov extension theorem [3–6] defines the minimal properties finite probability distributions have to satisfy in order for an underlying process to exist:

Theorem. [Kolmogorov] *Let Λ be a set of times. For each finite $\Lambda_k \subseteq \Lambda$, let $\mathbb{P}_{\Lambda_k}$ be a (sufficiently regular) k-step joint probability distribution. There exists an underlying stochastic process $\mathbb{P}_\Lambda$ that satisfies $\mathbb{P}_{\Lambda_k} = \mathbb{P}_\Lambda^{|\Lambda_k}$ for all finite $\Lambda_k \subseteq \Lambda$ iff $\mathbb{P}_{\Lambda_k} = \mathbb{P}_{\Lambda_\ell}^{|\Lambda_k}$ for all $\Lambda_k \subseteq \Lambda_\ell \subseteq \Lambda$.*

In other words, if a family of joint probability distributions on finite sets of times satisfies a consistency condition (as well as an additional minor regularity property [5, 6]) there is an underlying stochastic process on Λ that has the distributions $\{\mathbb{P}_{\Lambda_k}\}_{\Lambda_k \subset \Lambda}$ as marginals. As stated above, the KET defines the notion of a classical stochastic process and reconciles the existence of an underlying process with its manifestations for finite times. It also enables the modelling of stochastic processes: Any mechanism that leads to finite joint probability distributions that satisfy a consistency condition is ensured to have an underlying process. For example, the proof of the existence of Brownian motion relies on the KET as a fundamental ingredient [19–22].

We emphasize that, in the (physically relevant) case where Λ is an infinite set, the probability distribution $\mathbb{P}_\Lambda$ can generally not be experimentally accessed. For example, in the case of Brownian motion, the set Λ could contain all times in the interval $[0, t]$ and each realization $\mathbf{i}_\Lambda$ would represent a possible continuous trajectory of a particle over this time interval. While we assume the existence of these underlying trajectories (and hence the existence of $\mathbb{P}_\Lambda$) in experiments concerning Brownian motion, we often only access their finite time manifestations, *i.e.*, $\mathbb{P}_{\Lambda_k}$ for some Λ_k. The KET bridges the gap between the finite experimental reality and the underlying infinite stochastic process.

Loosely speaking, the KET holds for classical stochastic processes, because there is no difference between 'doing nothing' and conducting a measurement but 'not looking at the outcomes' (*i.e.*, summing over the outcomes at a time). Put differently, the validity of the KET is based on the fundamental assumption that the interrogation of a system does on average not influence its state. Consequently, marginalization is the correct way to obtain the descriptor for fewer times and any classical stochastic process leads to compatible finite joint probability distributions; this compatibility is independent of whether the system was interrogated or not, and the converse also holds. This fails to be true in causal modelling scenarios.

B. The KET and causal modelling

The compatibility of joint probability distributions for different sets of times hinges on the fact that observations in classical physics do not alter the state of the system that is being observed. In contrast to passive interrogations, that merely reveal information, active interventions, like they are used in the case of causal modelling, on average change the state of the interrogated system. Thus the future statistics after an intervention crucially depends on how the system was manipulated and the prerequisite of compatible joint probability distributions is generally not fulfilled anymore.

Consider, for example, the case of a pharmacologist that tries to understand the effect of different drugs they developed on a disease. In our simplified example, let the disease have two different symptoms S_a and S_b, and denote the absence of symptoms S_c. Whenever the pharmacologist observes S_a, they administer drug D_a, whenever they observe symptom S_b they administer drug D_b, and whenever they observe S_c they do nothing; this choice of actions defines an instrument $\mathcal{J}$. Running their trial with sufficiently many patients, the pharmacologist can deduce probability distributions for the occurrence of symptoms over time, given the drugs that were administered. For example, if the drugs were administered on three consecutive days, they would have obtained a probability distribution of the form $\mathbb{P}_{\Lambda_3}(s_3, s_2, s_1|\mathcal{J}_3 = \mathcal{J}, \mathcal{J}_2 = \mathcal{J}, \mathcal{J}_1 = \mathcal{J})$, where $s_\alpha \in \{S_a, S_b, S_c\}$, and the instruments (*i.e.*, the drug administration rule) are the same each day. However, this data would not allow them to find out what would have happened, had they *not* administered drugs on day two, *i.e.*, $\sum_{t_2} \mathbb{P}_{\Lambda_3}(s_{t_3}, s_{t_2}, s_{t_1}|\mathcal{J}_3, \mathcal{J}_2, \mathcal{J}_1) \neq \mathbb{P}_{\Lambda_3\setminus\{t_2\}}(s_{t_3}, s_{t_1}|\mathcal{J}_3, \mathcal{J}_1)$; intermediate interventions change the state of the interrogated system, and hence the future statistics that are being observed; for another, more numerically tangible example, see Fig. 2. Consequently, probability distributions do generally not satisfy compatibility conditions when interventions are allowed.

Compatibility can fail to hold whenever the system of interest is actively interrogated. In particular, it fails to hold in quantum mechanics, where even projective measurements in general change the state of a system on average, and interventions are not just an experimental choice but unavoidable.

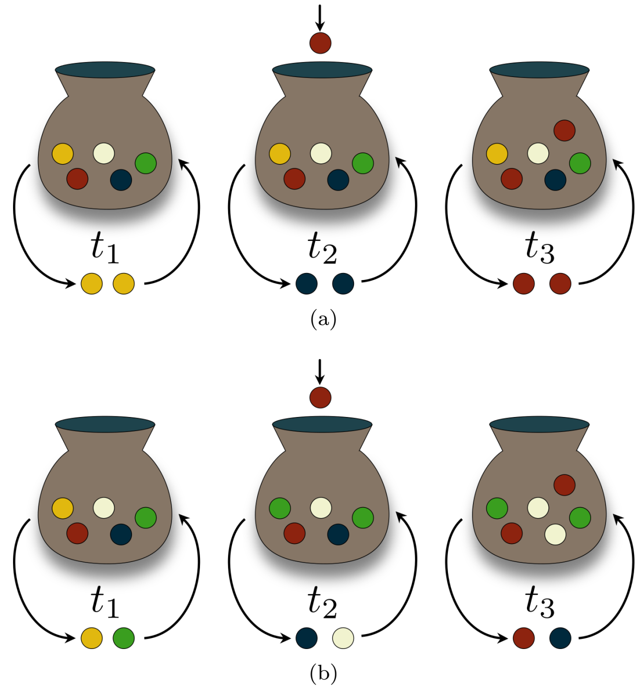

Figure 2. **(a)** *Classical process without intervention.* A possible three step process is the drawing with replacement of colored balls from an urn. In our example, independent of the actions of the experimenter, a red ball drops into the urn at t_2 (this could, *e.g.*, represent the interaction with an uncontrollable environment.) The experimenter can deduce the joint probability distribution $\mathbb{P}_{\{t_3,t_2,t_1\}}(c_3, c_2, c_1)$, to draw different sequences of colors. $\mathbb{P}_{\{t_3,t_2,t_1\}}$ contains all distributions for fewer times, for example $\mathbb{P}_{\{t_3,t_1\}}, \mathbb{P}_{\{t_3,t_2\}}$, and $\mathbb{P}_{\{t_1\}}$. **(b)** *Classical process with intervention.* Instead of putting the same ball back in the urn, the experimenter could exchange it with a different color (for example, upon drawing yellow, they could replace it with green at t_1, replace blue with white at t_2 and replace red with blue at t_3). The respective replacement rules are encapsulated in the instruments $\mathcal{J}_3$, $\mathcal{J}_2$, and $\mathcal{J}_1$. Now, from the probability distribution $\mathbb{P}_{\{t_3,t_2,t_1\}}(c_3, c_2, c_1|\mathcal{J}_3, \mathcal{J}_2, \mathcal{J}_1)$, it is generally not possible to deduce probability distributions for fewer steps, like, *e.g.*, $\mathbb{P}_{\{t_3,t_1\}}(c_3, c_1|\mathcal{J}_3, \mathcal{J}_1)$, or $\mathbb{P}_{\{t_3,t_2\}}(c_3, c_2|\mathcal{J}_3, \mathcal{J}_2)$. This lack of consistency can not be remedied by simple relabeling of the times due to the red ball that drops into the urn at t_2. Note that if all instruments are the idle instrument, the case with interventions coincides with the case without interventions.

IV. (QUANTUM) STOCHASTIC PROCESSES WITH INTERVENTIONS

A. The KET in QM

As hinted at throughout this work, descriptions of quantum mechanical processes must necessarily account for the fundamental invasiveness of measurements, which renders the KET invalid for the same reason that some choices of intervention do in the case of classical causal modelling. To see how even projective measurements in quantum mechanics lead to families of probability distributions that do not satisfy the KET, consider the following concatenated Stern-Gerlach experiment: Let the initial state of a spin-$\frac{1}{2}$ particle be $|+\rangle = \frac{1}{\sqrt{2}}(|{\uparrow}\rangle + |{\downarrow}\rangle)$, where $|{\uparrow}\rangle$ and $|{\downarrow}\rangle$ are the spin-up and spin-down state in the z-direction, respectively. Now, we measure the state sequentially in the z-, x- and z-directions at times t_1, t_2 and t_3. These measurements have the possible outcomes $\{\uparrow, \downarrow\}$ and $\{\rightarrow, \leftarrow\}$ for the measurement in z- and x-direction, respectively. It is easy to see that the probability for any possible sequence of outcomes is equal to $1/8$. For example, we have

$$\mathbb{P}_{\Lambda_3}(\uparrow, \rightarrow, \uparrow|\mathcal{J}_z, \mathcal{J}_x, \mathcal{J}_z) = \mathbb{P}_{\Lambda_3}(\uparrow, \leftarrow, \uparrow|\mathcal{J}_z, \mathcal{J}_x, \mathcal{J}_z) = \frac{1}{8},$$

where $\mathcal{J}_z$ and $\mathcal{J}_x$ represent the instruments used to measure in the z- and x- direction respectively, and

$\Lambda_3 = \{t_3, t_2, t_1\}$. Summing over the outcomes at time t_2, we obtain the marginal probability $\mathbb{P}^{|\{t_1,t_3\}}_{\Lambda_3}(\uparrow,\uparrow|\mathcal{J}_z,\mathcal{J}_z) = 1/4$. However, by considering the case where the measurement is not made at t_2, it is easy to see that $\mathbb{P}_{\{t_3,t_1\}}(\uparrow,\uparrow|\mathcal{J}_z,\mathcal{J}_z) = 1/2$. The intermediate measurement changes the state of the system, and the corresponding probability distributions for different sets of times are not compatible anymore [16, 23].

It is important to highlight the close relation of this breakdown of consistency and the violation of Leggett-Garg inequalities in quantum mechanics [7, 9]. The assumption of consistency between descriptors for different sets of times that is crucial for the derivation of the KET subsumes the assumptions of macroscopic realism and noninvasive measurability that are the basic principles leading to the derivation of Leggett-Garg inequalities: While macroscopic realism implies that joint probability distributions for a set of times can be expressed as marginals of a joint probability distribution for more times, non-invasiveness means that all finite distributions are marginals of the *same* distribution. For example, the two-step joint probability distributions $\mathbb{P}_{\{t_2,t_1\}}$, $\mathbb{P}_{\{t_3,t_2\}}$, and $\mathbb{P}_{\{t_3,t_1\}}$, that are considered in the Leggett-Garg scenario are all marginals of a three-step distribution $\mathbb{P}_{\{t_3,t_2,t_1\}}$. As soon as non-invasiveness and/or macrorealism don't hold anymore, the KET can fail and Legget-Garg inequalities can be violated.

Nevertheless, there should be some compatibility between descriptors for different sets of times; the breakdown of the KET should be a problem of the formalism rather than a physical fact. We now show that a change of perspective enables one to prove an extension theorem in quantum mechanics and any theory with interventions.

B. Generalized extension theorem (GET)

Processes involving interventions, including quantum processes and those in classical causal modelling, do not lead to compatible joint probability distributions for different sets of times in general. This problem can be remedied by assuming the standpoint of quantum causal modelling, and choosing a description of such stochastic processes that takes interventions and their corresponding change of the system into account. With this description, it is possible to recover a compatibility property that is satisfied by any process with interventions, and a generalized extension theorem can be derived.

As in the classical causal modelling case, in quantum mechanics, an experimenter can choose an instrument $\mathcal{J}_a$ at each time t_a, and every outcome i_a corresponds to a particular transformation of the system that is interrogated. Mathematically, an observation of outcome i_a given the instrument $\mathcal{J}_a$ corresponds to a completely positive (CP) map $\mathcal{M}_{i_a}$ that describes the change of the state of the system [24, 25]. The set of possible CP maps comprising an instrument add up to a completely positive trace preserving (CPTP) map $\sum_{i_\alpha}\mathcal{M}_{i_\alpha}$, which describes the overall average transformation applied by the instrument. While the introduction of maps (or, more generally, *events* [23]) in classical physics (without interventions) is superfluous, it is fundamentally unavoidable in quantum mechanics, as well as in more general probabilistic theories [23, 26]; every measurement alters the state of the system of interest, and a full description of a temporal process necessitates knowledge of how the system is changed at each time.

As in the example from Sec. IV A, an experimenter could choose to measure a system in different bases. The projective measurement in the basis $\{|i\rangle\}$ at t_α of a state ρ that yields outcome i_α would be described by a CP map of the form $\mathcal{M}_{i_\alpha}[\rho] = \langle i_\alpha|\,\rho\,|i_\alpha\rangle\,|i_\alpha\rangle\langle i_\alpha|$, where $|i_\alpha\rangle \in \{|i\rangle\}$. More generally, the measuring instrument need not preserve the measured state of the system, but could replace it entirely; upon measuring an outcome i_α (corresponding to a projection on a state $|i_\alpha\rangle$), a different instrument could leave the system in the state ρ_{i_α}, with a resulting CP map $\mathcal{M}_{i_\alpha}[\rho] = \langle i_\alpha|\,\rho\,|i_\alpha\rangle\,\rho_{i_\alpha}$. In the most general case, the experimenter could perform *any* (trace non-increasing) CP map, including deterministic operations, such as unitary transformations. We will employ the convention that, for a given instrument $\mathcal{J}_\alpha$, the CP map corresponding to the outcome i_α is denoted by $\mathcal{M}_{i_\alpha}$.

In this language, each realization of an experiment corresponds to a sequence of CP maps that transform the system at a series of times, and the set of possible CP maps that could be applied is dictated by the choice of instruments used to interrogate the system in question. A quantum process is fully characterized once *all* of the probabilities $\mathbb{P}_{\Lambda_k}(i_k,\dots,i_1|\mathcal{J}_k,\dots,\mathcal{J}_1)$ for each such sequence with all possible instruments are known. Having all of these probability distributions at hand allows one to deduce the causal structure of a process, *i.e.*, it is the basis of quantum causal modelling [13, 15].

Written more succinctly, a k-step quantum process is fully characterized by an object $\mathcal{T}_{\Lambda_k}$ that maps sequences of CP maps to probabilities, *i.e.* $\mathcal{T}_{\Lambda_k}[\mathcal{M}_{i_k},\dots,\mathcal{M}_{i_1}]$ yields the probability $\mathbb{P}_{\Lambda_k}(i_k,\dots,i_1|\mathcal{J}_k,\dots,\mathcal{J}_1)$ to obtain the outcomes $i_k,\dots,i_1$ given the choices of instruments $\mathcal{J}_k,\dots,\mathcal{J}_1$ (see Fig. 3). In this sense, $\mathcal{T}_{\Lambda_k}$ represents a Born rule for temporal processes [27]. The mapping $\mathcal{T}_{\Lambda_k}$ is a completely positive multilinear functional that can be reconstructed in a finite number of experiments [12, 28–30]. Throughout the remainder of this Paper, we will call $\mathcal{T}_{\Lambda_k}$ a k-step *comb*, following Refs. [31–33].

A comb $\mathcal{T}_{\Lambda_k}$ contains all the multi-time correlations necessary to fully characterize a k-step quantum process. While the CP maps $\mathcal{M}_{i_\alpha}$ change the state of the system, they do not change the k-step process given by $\mathcal{T}_{\Lambda_k}$. Loosely speaking, the comb contains all parts of the dynamics that are not manipulated by the experimenter. This is analogous to the way in which the preparation of

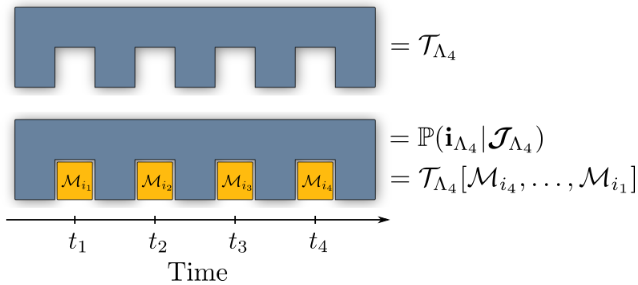

Figure 3. *Graphical representation of a four step quantum comb.* A four step comb can be represented as an object with four slots (each slot corresponds to a time $t_\alpha \in \Lambda_4$); it encodes all multi-time correlations between observables at those times. The probability to observe the outcomes $i_4, \dots, i_1$ given the instruments $\mathcal{J}_4, \dots, \mathcal{J}_1$ is obtained by inserting the corresponding CP maps into the comb, *i.e.*, by letting $\mathcal{T}_{\Lambda_4}$ act on $\mathcal{M}_{i_4}, \dots, \mathcal{M}_{i_1}$.

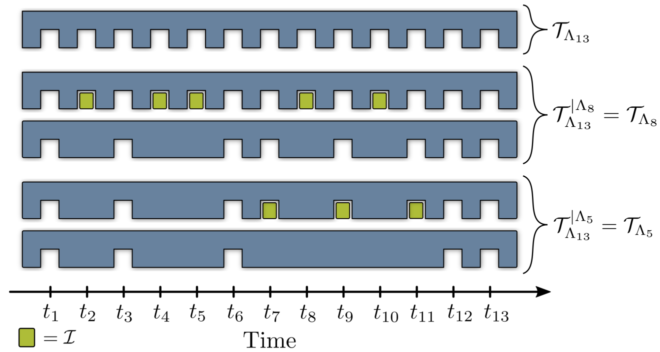

Figure 4. *Consistency condition for combs.* If there is an underlying process, any comb $\mathcal{T}_{\Lambda_k}$ can be obtained from $\mathcal{T}_{\Lambda_\ell}$ by letting $\mathcal{T}_{\Lambda_\ell}$ act on the identity map at the excessive times. Here, for the sets of times $\Lambda_{13} = \{t_{13}, \dots, t_1\}$, $\Lambda_8 = \{t_{13}, t_{12}, t_{11}, t_9, t_7, t_6, t_3, t_1\}$ and $\Lambda_5 = \{t_{13}, t_{12}, t_6, t_3, t_1\}$ we depict the containment of the comb $\mathcal{T}_{\Lambda_8}$ in $\mathcal{T}_{\Lambda_{13}}$ and the containment of $\mathcal{T}_{\Lambda_5}$ in both $\mathcal{T}_{\Lambda_{13}}$ and $\mathcal{T}_{\Lambda_8}$.

an initial state and the measurement of the final state in quantum process tomography do not influence the underlying dynamics (*i.e.* the CPTP map connecting input and output state).

Just as in the classical case, the knowledge of all relevant joint probability distributions (*i.e.*, the knowledge of $\mathcal{T}_{\Lambda_k}$) allows one to deduce causal relations between the k events in Λ_k. We emphasize that classical causal modelling is included in this quantum causal modelling framework as a special case. Whenever a system is measured and prepared in a fixed basis (using a classical instrument), and the process $\mathcal{T}_{\Lambda_k}$ also preserves this basis, the result is a set of joint distributions consistent with a classical causal model. From Proposition 1, it also follows that classical stochastic processes without interventions can be described by the same framework.

With this complete description on finite sets of time steps at hand, we can determine the compatibility condition between related processes. A family of combs that stems from an underlying (open) dynamics fulfills a natural consistency condition [29]; for any two sets of times $\Lambda_k \subseteq \Lambda_\ell$, the comb $\mathcal{T}_{\Lambda_k}$ can be obtained from $\mathcal{T}_{\Lambda_\ell}$ by letting it act on identity operations $\mathcal{I}_{t_\alpha}$ (with $\mathcal{I}_{t_\alpha}[\rho] = \rho$ for any state of the system ρ at time t_α) at times $t_\alpha \in \Lambda_\ell \setminus \Lambda_k$, *i.e.*,

$$\mathcal{T}_{\Lambda_k}[\,\cdot\,] = \mathcal{T}_{\Lambda_\ell}\left[\bigotimes_{t_\alpha \in \Lambda_\ell \setminus \Lambda_k} \mathcal{I}_{t_\alpha}, \,\cdot\,\right] := \mathcal{T}_{\Lambda_\ell}^{|\Lambda_k}[\,\cdot\,]\,, \qquad (2)$$

where we have employed the shorthand notation $\bigotimes_{t_\alpha \in \Lambda_\ell \setminus \Lambda_k} \mathcal{I}_{t_\alpha}$ to signify that the identity operation was 'implemented' at each time $t_\alpha \in \Lambda_\ell \setminus \Lambda_k$. The graphical representation of Eq. (2) is depicted in Fig. 4.

It is important to note the difference between Eq. (2) and the consistency condition for classical stochastic processes, stemming from the stronger notion of 'doing nothing' in the quantum case. If there is an underlying process, any comb can be obtained from one that applies to a larger set of times by letting it act on the identity map, which leaves any state unchanged, at the excessive times. This is by no means the same as computing the marginals of families of probability distributions that have been obtained for a fixed set of measurement instruments, which will only preserve states which are diagonal in a fixed basis. We recover descriptors for different sets of times that are compatible with each other only when we switch to a causal modelling description of the process. From this, we obtain our main result, the generalized extension theorem (GET):

Theorem (Generalized extension theorem)**.** *Let Λ be a set of times. For each finite $\Lambda_k \subseteq \Lambda$ let $\mathcal{T}_{\Lambda_k}$ be a k-step comb. There exists a general stochastic process $\mathcal{T}_\Lambda$ that satisfies $\mathcal{T}_{\Lambda_k} = \mathcal{T}_\Lambda^{|\Lambda_k}$, as defined in Eq.* (2)*, for all finite $\Lambda_k \subseteq \Lambda$ iff $\mathcal{T}_{\Lambda_k} = \mathcal{T}_{\Lambda_\ell}^{|\Lambda_k}$ for all $\Lambda_k \subseteq \Lambda_\ell \subseteq \Lambda$.*

The proof can be found in App. A. It proceeds analogously to that of the original Kolmogorov extension theorem, presented in [6]. The consistency property is used to uniquely define a comb $\mathcal{T}_\Lambda^\sharp$ on a 'sufficiently large' container space. $\mathcal{T}_\Lambda^\sharp$ can then be extended to a linear functional $\overline{\mathcal{T}}_\Lambda^\sharp$ that fulfills the properties of the comb $\mathcal{T}_\Lambda$ on the closure of said container space. As in the classical case [5], the underlying stochastic process characterized by $\mathcal{T}_\Lambda$ is – unlike $\overline{\mathcal{T}}_\Lambda^\sharp$ – not necessarily unique. Since the action of all possible $\mathcal{T}_\Lambda$ coincides with the unique $\overline{\mathcal{T}}_\Lambda^\sharp$ on a sufficiently large space, and hence yields the correct finite combs $\mathcal{T}_{\Lambda_k}$, this non-uniqueness cannot be detected experimentally and does not constitute a practical problem.

We emphasize that, even though we have phrased the above in the language of quantum mechanics, there is nothing particularly quantum mechanical about the GET. The proof of the theorem only uses the linearity and boundedness of the combs $\mathcal{T}_{\Lambda_k}$, as well as their compatibility. Consequently, it holds for *any* probabilistic theory (with interventions).

Furthermore, the input and output spaces of the CP maps the comb acts on do not have to be of the same dimension. In this case, the identity map used for the consistency condition has to be slightly generalized: A CPTP map $\mathcal{M}_{i_\alpha} : \mathcal{B}(\mathcal{H}_\alpha^{(1)}) \to \mathcal{B}(\mathcal{H}_\alpha^{(2)})$, where $\mathcal{B}(\mathcal{H}_\alpha^{(x)})$ is the space of bounded operators on the Hilbert space $\mathcal{H}_\alpha^{(x)}$, is implemented via a corresponding unitary U_{i_α}, a *fixed* ancillary state $\eta_\alpha \in \mathcal{B}(\mathcal{H}_{\mathrm{A}_\alpha})$, and a partial trace $\mathrm{tr}_{\mathrm{B}_\alpha}$ that is such that the resulting state $\mathcal{M}_{i_\alpha}[\rho] = \mathrm{tr}_{\mathrm{B}_\alpha}\left[U_{i_\alpha}(\rho \otimes \eta_\alpha) U_{i_\alpha}^\dagger\right]$ lies in $\mathcal{B}(\mathcal{H}_\alpha^{(2)})$. With this, we can define a *generalized identity map* $\mathcal{I}_{t_\alpha}^{(1\to 2)}[\rho] = \mathrm{tr}_{\mathrm{B}_\alpha}(\rho \otimes \eta_\alpha)$ and the GET still holds. Consequently, our theorem accounts for the case where particles are created/annihilated in the process, as well as the case where different degrees of freedom are manipulated at each time t_α, or where the number of measurement outcomes and active interventions differ.

In the derivation of the GET, we make the implicit assumption that probabilities only depend on the respective CP maps that where implemented, but not on the particular instrument that was used to implement them. This property has been dubbed 'instrument non-contextuality' [27, 34] or 'operational instrument equivalence' [23]. In principle, our derivation could be straightforwardly adapted to any theory, where this assumption is not satisfied anymore, but probabilities are still a linear function of the maps *and* their respective contexts (*i.e.*, the respective instrument). Instead of the identity map, one would then use a pair $(\mathcal{I}, \mathcal{J}_\mathcal{I})$ of identity map and identity context for marginalization, and the GET would still hold.

It is important to clearly delineate between the classical Kolmogorov extension theorem and the GET. The KET hinges on the fact that, in classical physics, a measurement does not change the average state of a system. This fails to hold in quantum mechanics, or any theory with interventions. In detail, in the language of quantum maps, the sum over the outcomes i_α of a measurement in a basis $\{|i\rangle\}$ at time t_α corresponds to the CPTP map $\mathcal{M}_\alpha = \sum_{i_\alpha} \mathcal{M}_{i_\alpha}$, where $\mathcal{M}_{i_\alpha}[\rho] = \langle i_\alpha | \rho | i_\alpha \rangle |i_\alpha\rangle\langle i_\alpha|$. In a classical stochastic process, the state ρ is diagonal in the basis $\{|i\rangle\}$, and we have $\mathcal{M}_\alpha[\rho] = \rho$; the average over measurement outcomes has the same effect as the classical 'do nothing' operation. As soon as ρ is not diagonal in the measurement basis, we have $\mathcal{M}_\alpha[\rho] \neq \rho$; on average, a measurement in quantum mechanics changes the state of the system and the future measurement statistics will depend on the measurement that was performed. Consequently, joint probability distributions in classical physics (without interventions) exhibit a consistency condition, while quantum processes (and theories with interventions) generally do not.

As in the classical case, the proof of the GET does not assume an *a priori* temporal ordering. The sets Λ_k could be sets of times, but also labels of different laboratories. We have the following remark:

Remark. *The proof of the GET does not assume any ordering of the sets* Λ_k*, and only uses the generalized containment property* (2) *as its main ingredient.*

Consequently, the GET also applies to causally non-separable processes [35, 36], as the descriptors for different sets of laboratories would still satisfy a compatibility condition. However, these processes do not have a deterministic Stinespring dilation [37] and the interpretation of an underlying 'process' becomes much less clear in the absence of a definitive causal ordering. We will briefly remark on this further in our conclusions, but leave a full exploration of this interpretation as an open question for future work. Next, we will see that the distinction between stochastic processes and causal modelling does not exist in the general case.

C. Quantum stochastic processes and quantum causal modelling

Using an instrument at some intermediate time t_α alters the state of a quantum system (even when averaging over all outcomes) and influences the statistics of later measurements in a non-negligible way. Nevertheless, the *full* descriptor of an ℓ-step process, *i.e.*, $\mathcal{T}_{\Lambda_\ell}$, contains all descriptors $\mathcal{T}_{\Lambda_k}$ for fewer times $\Lambda_k \subseteq \Lambda_\ell$, and a family of compatible combs implies the existence of an underlying stochastic process $\mathcal{T}_\Lambda$.

Like in the classical case, the GET defines the notion of an underlying stochastic process in quantum mechanics, or any other theory with interventions, and fixes the minimal necessary requirements for the existence of an underlying process. As we have seen, in quantum mechanics, it is unavoidable to employ a description that takes interventions into account, when attempting to obtain a consistent description of a quantum process; if one wants to properly define quantum stochastic processes, one is forced to use the framework of causal modelling. This motivates the following proposition:

Proposition 2. *The theory of quantum causal modelling and the theory of quantum stochastic processes are equivalent.*

In contrast to Proposition 1, the set of quantum causal models does not just contain the set of quantum stochastic processes but coincides with it; in classical physics, we obtain a consistent description of stochastic processes without taking interventions into account, and we can *choose* to intervene if we want to probe the causal structure of a process. In quantum mechanics, a consistent description of stochastic processes can *only* be recovered if interventions are included in the description from the start. Interventions are not a choice but a necessity in quantum mechanics, which leads to the equivalence of quantum causal modelling and quantum stochastic processes.

This implies that the breakdown of the KET in quantum mechanics is *fundamental*, while it can in principle be removed by changing perspective in a classical process with interventions. In the latter case, a *super-observer*, that observes both the experimenter manipulating the system of interest as well as the stochastic process itself, would obtain families of joint probability distributions that display a compatibility property. Put differently, for classical processes, by incorporating the experimenter and their choice of instrument into the stochastic process, the KET always applies on a higher level. In quantum mechanics, this is generally not true. No matter the level at which a super-observer observes a process, the respective joint probability distributions do not satisfy a compatibility property, and the KET fails to hold. This fundamental breakdown of the KET in quantum mechanics is mirrored by no-go theorems that show that non-contextual theories cannot reproduce the predictions of quantum mechanics; for many of these theorems, the notion of ontic latent variables [38, 39] or ontic processes [23] are introduced, and the basic assumption is made that the distributions over observable outcomes can be obtained by marginalization of a larger joint distribution over the values of the ontic variable. Subsequently, it is shown that, together with other assumptions, this prerequisite fails to reproduce predictions made by quantum mechanics. The GET dictates how to correctly compute marginals in quantum mechanics, such that all resulting probability distributions 'fit together' and are the marginals of one common comb $\mathcal{T}_\Lambda$. It is therefore conceivable that a derivation starting from the assumption of compatibility in the sense of the GET would lead to theories that can indeed reproduce quantum mechanics.

We reiterate that classical stochastic processes are a very special subset of general stochastic processes, namely the ones where the system of interest is never rotated out of its fixed (pointer) basis, and the experimenter can only perform projective measurements in this basis. We now show that the KET can be derived in a straight forward way as a corollary of the GET.

D. GET $\Rightarrow$ KET

Our generalised extension theorem applies to a strictly larger class of theories than the standard KET and includes the latter as a corollary. We have the following proposition:

Proposition 3. *The GET implies the KET.*

The detailed proof of this statement can be found in App. B. There, we show that a family of compatible probability distributions $\mathbb{P}_{\Lambda_k}$ can be mapped onto a family of combs $\mathcal{T}^{\mathrm{cl}}_{\Lambda_k}$ that satisfy the consistency condition of the GET. The existence of the underlying process $\mathcal{T}_\Lambda$ then ensures the existence of a joint probability distribution $\mathbb{P}_\Lambda$ that has all finite ones as marginals.

While the original version of the KET does not hold for quantum processes, it is important to note that the breakdown of the compatibility property of joint probability distributions is not a signature of quantum mechanics *per se*; as we have already seen, any framework that allows for interventions will exhibit this feature. The GET provides a proper theoretical underpinning for the corresponding experimental situations. On the other hand, the breakdown of the compatibility property can happen in quantum mechanics even if only projective measurements in a fixed basis $\{|i_\alpha\rangle\}$ are performed [16, 17].

As already mentioned, the absence of compatibility is tantamount to the absence of either macrorealism, or non-invasiveness (or both). Consequently, it can be used as a definition of non-classicality, as proposed in Ref. [17]. There, the authors employ the breakdown of the consistency condition on the level of probability distributions, when measuring in a fixed basis, as a means to define the non-classicality of Markovian processes. Using the framework of quantum combs for the description of quantum stochastic processes the ideas of [17] can be extended to general processes with memory, *i.e.*, non-Markovian processes.

Following Ref. [17], we consider an ℓ-step process to be classical if its joint probability distributions with respect to measurements in a fixed basis $\{|i_\alpha\rangle\}$ satisfy a consistency condition. Put differently, an ℓ-step process $\mathcal{T}_{\Lambda_k}$ is classical (with respect to the basis $\{|i\rangle\}$) iff for all $\Lambda_k \subseteq \Lambda_\ell$ and all possible sequences of outcomes $i_k, \dots, i_1$

$$\mathcal{T}_{\Lambda_k}[\mathcal{P}_{i_k}, \dots, \mathcal{P}_{i_1}] = \sum_{\Lambda_\ell \setminus \Lambda_k} \mathcal{T}_{\Lambda_\ell}[\mathcal{P}_{i_\ell}, \dots, \mathcal{P}_{i_1}], \tag{3}$$

where $\mathcal{P}_{i_\alpha}$ corresponds to obtaining outcome i_α from a projective measurement in a fixed basis at time t_α, *i.e.*, $\mathcal{P}_{i_\alpha}[\rho] = \langle i_\alpha|\,\rho\,|i_\alpha\rangle\,|i_\alpha\rangle\langle i_\alpha|$.

We provide an example of classical combs that satisfy Eq. (3) in our proof of Proposition 3 (specifically in Eq. (B3) of Appendix B). The general structure of classical combs that satisfy Eq. (3) can be analyzed using the Choi isomorphism between quantum processes and positive matrices (again, see App. B). As combs can describe general processes with memory, Eq. (3) represents the consistent definition of classical processes with memory and allows a direct extension of the results obtained in Ref. [17] to the non-Markovian case.

V. RELATION TO PREVIOUS WORKS

As already mentioned, the proof of the GET does not rely on any particularities that are exclusive to quantum mechanics or our formulation thereof. The GET constitutes a sound basis for the description of *any* conceivable (classical, quantum or beyond) theory of stochastic processes with interventions – independent of the employed framework.

While we referred throughout to the framework of *quantum combs* [31–33], originally derived as the most general representation of quantum circuit architectures, our results apply equally well to any other framework for describing quantum processes as linear functionals. Examples of the mathematical objects and frameworks (often the same thing under a different name) given a firm theoretical foundation by the GET include: *process tensors* [12, 28, 29] and *causal automata/non-anticipatory channels* [40, 41], which describe the most general open quantum processes with memory; *causal boxes* [42] that enter into quantum networks with modular elements; *operator tensors* [43, 44] and *superdensity matrices* [45], employed to investigate quantum information in general relativistic space-time; and, finally, *process matrices*, used for quantum causal modelling [13–15, 35]. In classical physics, as well as the standard causal modelling framework discussed in Sec. II, our result applies to the ϵ-transducers used within the framework of computational mechanics [46, 47] to describe processes with active interventions.

Our theorem proves the existence of a container space for all of the aforementioned frameworks and allows for their complete and consistent representation in the continuous time limit, thus providing an overarching theorem for probabilistic theories with interventions. This is of particular importance for the field of open quantum mechanics where the lack of an extension theorem has been a roadblock to obtaining a framework that coincides with classical descriptions in the correct limit [16]. Here, switching perspective allows one to describe both classical as well as quantum open systems in a unified framework. This fact has recently been used to obtain an unambiguous definition of non-Markovianity in quantum mechanics that coincides with the classical one in the correct limit [48].

The GET goes beyond previous attempts to generalize the KET for quantum mechanics. An extension theorem for positive operator valued measures was derived in Ref. [49] and was used in Ref. [50] to show the existence of an 'infinite composition' of an instrument. This extension theorem is, however, limited to particular cases of positive operator valued measures, and not general enough to provide an underpinning for the description of stochastic processes with interventions.

More generally, a version of the KET for quantum processes was derived in Ref. [18]. In this work, the authors showed that any quantum stochastic "process can be reconstructed up to equivalence from a projective family of correlation kernels". By decomposing the control operations $\mathcal{M}_{i_\alpha}$ into their component Kraus operators, it can explicitly be shown that these correlation kernels correspond to combs, and consequently, for quantum processes, the GET is equivalent to Thm. 1.3 in Ref. [18]. However, the mathematical structure of the latter does not tie in easily with recently developed frameworks for the description of quantum (or classical) causal modelling, nor does it lend itself in a straightforward way to the discussion of their key properties. The structural features of combs render the investigation of fundamental features of a process, like their non-Markovianity [28, 48], their causal structure [13, 35, 42], and their classicality tractable.

Our formulation has the advantage that combs are defined in a clear-cut *operational* way, and allow for a generalized Stinespring dilation [29, 33], which makes their interpretation in terms of open quantum system dynamics straightforward. Additionally, even though the GET is stated for combs that map sequences of CP maps to probabilities, its proof also applies – with slight modifications – to general quantum combs (*i.e.*, maps that map combs onto combs [32, 33]).

VI. CONCLUSIONS

While the KET is the fundamental building block for the theory of classical stochastic processes, it does not hold in quantum mechanics, or any other theory that allows for active interventions. This breakdown goes hand in hand with the violation of Leggett-Garg inequalities: the violation of such an inequality *always* implies that compatibility conditions are not satisfied, and hence the KET does not hold.

In this work, we have proven a generalized extension theorem that applies to any process with interventions, including quantum ones. We have therefore shown that the roadblocks encountered when describing quantum processes in terms of joint probability distributions can be remedied by changing perspective; while the evolution of a density matrix over time does not contain enough statistical information for consistency properties to hold [16], considering a quantum stochastic process as a linear functional acting on sequences of CP maps allows one to formulate a fully fledged theory. Taking interventions into account is the only way to obtain a consistent definition and rigorous mathematical foundation for quantum stochastic processes. Put differently, without taking interventions into account, there is no way to consistently define quantum stochastic processes. In this sense, two seemingly different frameworks – the framework of causal modelling, and the theory of quantum stochastic processes – are actually two sides of the same coin.

In the limit of continuous time, the sequence of CP maps becomes a continuous driving/control of the system of interest. Thus, the GET provides the theoretical foundation for these experimental scenarios, important for development of quantum technologies. Likewise, just as in the case of classical stochastic processes, the GET provides a toolbox for the modelling of quantum stochastic processes; any mechanism that leads to consistent families of combs automatically defines an underlying process.

It is important to emphasize the generality of our main result. Due to the linearity of mixing, *any* meaningful description of a stochastic process – quantum or not – must be expressible in terms of a linear function on the space of locally accessible operations [12]. The proof of the GET is versatile enough to account for any framework that aims to describe temporally ordered processes, and hence provides a sound mathematical underpinning for all of them.

The GET contains the original KET as the special case where the family of processes is diagonal in the reference basis, and the only allowed CP maps are projective measurements in the same basis. On the one hand, this implies that our extension of classical processes to the quantum realm is the correct one. On the other hand, this clear-cut definition of classical combs lends itself ideally to the investigation of the interplay of coherence and classicality, as proposed in Ref. [17], in the experimental observation of real-world processes with memory.

Finally, while we have mostly discussed temporally ordered processes, in principle, even causally disordered processes could be described by families of functionals that satisfy a consistency requirement (Λ would then be thought of as a set of labels for different laboratories). However, there is no deterministic Stinespring dilation for causally disordered processes [37]. There are, on the other hand, dilations that include post-selection [31, 51], and we conjecture that an underlying causally disordered stochastic process would be equivalent to post-selection on a class of trajectories resulting from continuous weak measurement.

ACKNOWLEDGMENTS

We thank Andrea Collevecchio, Tim Garoni and Top Notoh for valuable discussions, and Roderich Tumulka for clarifying remarks on the KET for POVMs. KM and FAP thank an anonymous referee for pointing them to the lack of a KET in quantum mechanics. SM is supported by the Monash Graduate Scholarship (MGS), Monash International Postgraduate Research Scholarship (MIPRS) and the J L William Scholarship. FS is grateful to thank Sri-Trang Thong Scholarship, Faculty of Science and "Postgraduate Student Exchange Program 2016", International Relations Division, Mahidol University, for financial support during his visit to Monash University, Australia. KM is supported through ARC FT160100073.

Appendix A: Proof of the Generalized extension theorem (GET)

Here, we prove the general extension theorem for processes with interventions. The structure of the proof follows the derivation of the KET presented in [6]; given a family of compatible combs, we use the consistency condition to define a unique comb $\mathcal{T}_\Lambda^\sharp$, that contains all finite ones as 'marginals', on a large enough 'container space'. Due to its properties $\mathcal{T}_\Lambda^\sharp$ can then be uniquely extended to a comb $\overline{\mathcal{T}}_\Lambda^\sharp$ on the closure of said container space.

Let Λ be a (possibly uncountable) set, $\{\Lambda_k\}_{\Lambda_k\subseteq\Lambda}$ the set of all finite subsets of Λ, and let $\{\mathcal{T}_{\Lambda_k}\}_{\Lambda_k\subseteq\Lambda}$ be the corresponding family of combs. For ease of notation, we assume all CP maps that the combs act on to have the same input and output space, *i.e.*, $\mathcal{M}_{i_\alpha}:\mathcal{B}(\mathcal{H}_\alpha)\to\mathcal{B}(\mathcal{H}_\alpha)$; a generalization to maps with distinct input and output spaces is straightforward. We denote the space of these maps by $\mathcal{L}(\mathcal{H}_\alpha)$. Consequently, we have $\mathcal{T}_{\Lambda_k}:\mathcal{L}_{\Lambda_k}\to\mathbb{R}$, where $\mathcal{L}_{\Lambda_k}=\bigotimes_{\alpha\in\Lambda_k}\mathcal{L}(\mathcal{H}_\alpha)$ and $\bigotimes_{\alpha\in\Lambda_k}$ denotes a tensor product over all times $t_\alpha\in\Lambda_k$. Let the family of combs satisfy the consistency condition $\mathcal{T}_{\Lambda_k}[\cdot]=\mathcal{T}_{\Lambda_\ell}\left[\bigotimes_{t_\alpha\in\Lambda_\ell\setminus\Lambda_k}\mathcal{I}_{t_\alpha},\cdot\right]$ for all finite $\Lambda_k\subseteq\Lambda_\ell\subseteq\Lambda$.

Now, we 'lift' the family of combs to a comb $\mathcal{T}_\Lambda^\sharp$ that contains all of them as 'marginals'. To this end, we define the inverse projection $\pi_{\Lambda_k}^{-1}:\mathcal{L}_{\Lambda_k}\to\mathcal{L}_\Lambda$, with $\pi_{\Lambda_k}^{-1}[\xi_{\Lambda_k}]=\xi_{\Lambda_k}\bigotimes_{t_\alpha\in\Lambda\setminus\Lambda_k}\mathcal{I}_{t_\alpha}$ for all $\xi_{\Lambda_k}\in\mathcal{L}_{\Lambda_k}$, which trivially extends any ξ_{Λ_k} to a corresponding operator that lies in $\mathcal{L}_\Lambda$. The operator $\pi_{\Lambda_k}^{-1}[\xi_{\Lambda_k}]$ exists and is unique for any finite $\Lambda_k\in\Lambda$ and all $\xi_{\Lambda_k}\in\mathcal{L}_{\Lambda_k}$ [52]. In the same way, we can define a partial inverse projection $\pi_{\Lambda_k\leftarrow\Lambda_\ell}^{-1}:\mathcal{L}_{\Lambda_k}\to\mathcal{L}_{\Lambda_\ell}$ for any two finite sets $\Lambda_k\subseteq\Lambda_\ell\subseteq\Lambda$, *i.e.*, $\pi_{\Lambda_k\leftarrow\Lambda_\ell}^{-1}[\xi_{\Lambda_k}]=\xi_{\Lambda_k}\bigotimes_{t_\alpha\in\Lambda_\ell\setminus\Lambda_k}\mathcal{I}_{t_\alpha}$. In terms of partial inverse projections, the consistency property reads $\mathcal{T}_{\Lambda_k}[\xi_{\Lambda_k}]=\mathcal{T}_{\Lambda_\ell}\left[\pi_{\Lambda_k\leftarrow\Lambda_\ell}^{-1}[\xi_{\Lambda_k}]\right]$.

Let $\mathcal{L}_\Lambda^\sharp$ denote the set of all 'lifted' operators, *i.e.*, $\mathcal{L}_\Lambda^\sharp=\left\{\xi\in\mathcal{L}_\Lambda:\xi=\xi_{\Lambda_k}\bigotimes_{t_\alpha\in\Lambda\setminus\Lambda_k}\mathcal{I}_{t_\alpha},\text{for some finite }\Lambda_k\right\}$. It is straightforward to see that this set forms a vector space; for any $\alpha,\beta\in\mathbb{R}$ and $\xi=\pi_{\Lambda_k}^{-1}[\xi_{\Lambda_k}],\zeta=\pi_{\Lambda_\ell}^{-1}[\zeta_{\Lambda_\ell}]$, we have $\alpha\,\xi=\pi_{\Lambda_k}^{-1}[\alpha\,\xi_{\Lambda_k}]\in\mathcal{L}_\Lambda^\sharp$ and $\alpha\,\xi+\beta\,\zeta=\pi_{\Lambda_k\cup\Lambda_\ell}^{-1}[\Gamma_{\Lambda_k\cup\Lambda_\ell}]\in\mathcal{L}_\Lambda^\sharp$, where $\Gamma_{\Lambda_k\cup\Lambda_\ell}=\pi_{\Lambda_k\leftarrow\Lambda_k\cup\Lambda_\ell}^{-1}[\alpha\,\xi_{\Lambda_k}]+\pi_{\Lambda_\ell\leftarrow\Lambda_k\cup\Lambda_\ell}^{-1}[\beta\,\zeta_{\Lambda_\ell}]\in\mathcal{L}_{\Lambda_k\cup\Lambda_\ell}$. Additionally, $\mathcal{L}_\Lambda^\sharp$ becomes a normed vector space by setting $\|\xi\|=\|\pi_{\Lambda_k}^{-1}[\xi_{\Lambda_k}]\|:=\|\xi_{\Lambda_k}\|_{\mathrm{op}}$, where $\|\cdot\|_{\mathrm{op}}$ is the operator norm in $\mathcal{L}_{\Lambda_k}$. On this vector space, we can define the comb $\mathcal{T}_\Lambda^\sharp$ via $\mathcal{T}_\Lambda^\sharp[\xi]:=\mathcal{T}_{\Lambda_k}[\xi_{\Lambda_k}]$, where $\xi=\pi_{\Lambda_k}^{-1}[\xi_{\Lambda_k}]$. $\mathcal{T}_\Lambda^\sharp$ is well-defined: if there are two different operators $\xi_{\Lambda_k}\in\mathcal{L}_{\Lambda_k}$ and $\xi_{\Lambda_\ell}\in\mathcal{L}_{\Lambda_\ell}$, such that $\xi=\pi_{\Lambda_k}^{-1}[\xi_{\Lambda_k}]=\pi_{\Lambda_\ell}^{-1}[\xi_{\Lambda_\ell}]$, the consistency property ensures that $\mathcal{T}_\Lambda^\sharp[\xi]$ is unique; it is straightforward to see that $\pi_{\Lambda_k\leftarrow\Lambda_k\cup\Lambda_\ell}^{-1}[\xi_{\Lambda_k}]=\pi_{\Lambda_\ell\leftarrow\Lambda_k\cup\Lambda_\ell}^{-1}[\xi_{\Lambda_\ell}]:=$

$\xi_{\Lambda_k \cup \Lambda_\ell}$. Employing the consistency condition yields $\mathcal{T}_{\Lambda_k}[\xi_{\Lambda_k}] = \mathcal{T}_{\Lambda_k\cup\Lambda_\ell}[\pi^{-1}_{\Lambda_k\leftarrow\Lambda_k\cup\Lambda_\ell}[\xi_{\Lambda_k}]] = \mathcal{T}_{\Lambda_k\cup\Lambda_\ell}[\xi_{\Lambda_k\cup\Lambda_\ell}] = \mathcal{T}_{\Lambda_k\cup\Lambda_\ell}[\pi^{-1}_{\Lambda_\ell\leftarrow\Lambda_k\cup\Lambda_\ell}(\xi_{\Lambda_\ell})] = \mathcal{T}_{\Lambda_\ell}[\xi_{\Lambda_\ell}]$, and consequently $\mathcal{T}^\sharp_\Lambda[\xi]$ is independent of the representation of ξ.

The operator $\mathcal{T}^\sharp_\Lambda$ is bounded, because every $\mathcal{T}_{\Lambda_k}$ is bounded. It is also linear; due to the linearity of $\mathcal{T}_{\Lambda_k}$ and the linearity of the inverse projection operators, we have $\mathcal{T}^\sharp_\Lambda[\alpha\xi+\beta\eta] = \alpha\mathcal{T}^\sharp_\Lambda[\xi] + \beta\mathcal{T}^\sharp_\Lambda[\eta]$ for all $\alpha,\beta\in\mathbb{R}$ and $\eta,\xi\in\mathcal{L}^\sharp_\Lambda$. Any linear bounded transformation from a normed vector space X to a normed complete vector space Y can be uniquely extended to a linear transformation from the completion $\overline{X}$ of X to Y [53]. Consequently, there exists a unique comb $\overline{\mathcal{T}}^\sharp_\Lambda$ defined on the completion $\overline{\mathcal{L}}^\sharp_\Lambda$ of $\mathcal{L}^\sharp_\Lambda$ (this completion is sometimes called *quasilocal algebra* in the literature [40]) that has – by construction – the family $\{\mathcal{T}_{\Lambda_k}\}_{\Lambda_k\subseteq\Lambda}$ as 'marginals'. This concludes the proof. □

It is important to note that $\overline{\mathcal{L}}^\sharp_\Lambda$ does *not* coincide with $\mathcal{L}_\Lambda$ (they coincide iff Λ is finite [52]). Consequently, there might be different combs $\mathcal{T}_\Lambda$ defined on $\mathcal{L}_\Lambda$ with coinciding action on all elements of $\overline{\mathcal{L}}^\sharp_\Lambda$. This, however, is not problematic. On the one hand, $\overline{\mathcal{L}}^\sharp_\Lambda$ "is in a way more important than" $\mathcal{L}_\Lambda$ because its elements arise from the ones of $\mathcal{L}(\mathcal{H}_\alpha)$ "by extension and algebraical and topological processes" [52]. On the other hand – just like for the KET [5] – the different possible combs on $\mathcal{L}_\Lambda$ all lead to the same measurement statistics on any experimentally accessible set of times, so this non-uniqueness is not accessible/detectable in practice.

Appendix B: Proof GET ⇒ KET

Here, we show that the GET contains the KET as a corollary. In detail, we show that any family of classical joint probability distributions can be mapped onto a family of quantum combs that satisfies the consistency condition. The GET then guarantees that there is an underlying classical comb $\mathcal{T}^{\text{cl}}_\Lambda$, and thus also an underlying classical process $\mathbb{P}_\Lambda$.

For the proof, we exploit the structural requirements a classical comb has to satisfy. To make these requirements evident, we employ the Choi-Jamiołkowski isomorphism (CJI) [54, 55] $M_{i_\alpha} = (\mathcal{M}_{i_\alpha}\otimes\mathcal{I})[|\Phi^+_\alpha\rangle\langle\Phi^+_\alpha|]$, that maps CP maps (or generally superoperators) $\mathcal{M}_{i_\alpha}:\mathcal{B}(\mathcal{H}^{(1)}_\alpha)\to\mathcal{B}(\mathcal{H}^{(2)}_\alpha)$ to bounded operators $M_{i_\alpha}\in\mathcal{B}(\mathcal{H}^{(2)}_\alpha\otimes\mathcal{H}^{(1)}_\alpha)$, by letting them act on one half of an unnormalized maximally entangled state $\Phi^+_\alpha := |\Phi^+_\alpha\rangle\langle\Phi^+_\alpha| = \sum_{i_\alpha,j_\alpha}^{d_\alpha}|i_\alpha i_\alpha\rangle\langle j_\alpha j_\alpha|\in\mathcal{B}(\mathcal{H}^{(1)}_\alpha\otimes\mathcal{H}^{(1)}_\alpha)$. For example, under this isomorphism, a projective measurement in the reference basis $\{|i_\alpha\rangle\}$ is mapped to an operator of the form $P_{i_\alpha} = |i_\alpha\rangle\langle i_\alpha|\otimes|i_\alpha\rangle\langle i_\alpha|$, while the 'do nothing' channel $\mathcal{I}_{t_\alpha}$ at time t_α is represented by Φ^+_α. Analogously, a comb can be mapped onto a (positive) matrix $\Upsilon_{\Lambda_k}\in\bigotimes_\alpha\mathcal{B}(\mathcal{H}^{(2)}_\alpha\otimes\mathcal{H}^{(1)}_\alpha)$ [29, 33]. In addition to positivity, the Choi matrices of a temporal process have to satisfy certain trace conditions to ensure proper causal ordering [32, 33]. All the combs we explicitly write down in what follows are constructed such that they automatically satisfy these conditions. Using the Choi matrix, the action of a comb can equivalently be expressed as

$$\mathcal{T}_{\Lambda_k}[\mathcal{M}_{i_k},\dots,\mathcal{M}_{i_k}] = \text{tr}\left[\left(M^{\text{T}}_{i_k}\otimes\cdots\otimes M^{\text{T}}_{i_1}\right)\Upsilon_{\Lambda_k}\right], \tag{B1}$$

where $\cdot^{\text{T}}$ denotes the transpose in the reference basis [12, 33].

In this notation, letting a comb act on identity maps amounts to projecting it on maximally entangled states, *i.e.*,

$$\Upsilon_{\Lambda_\ell|\Lambda_k} = \text{tr}_{\Lambda_\ell\setminus\Lambda_k}\left[\left(\mathbb{1}_{\Lambda_k}\bigotimes_{\alpha\in\Lambda_\ell\setminus\Lambda_k}\Phi^+_\alpha\right)\Upsilon_{\Lambda_k}\right], \tag{B2}$$

where $\text{tr}_{\Lambda_\ell\setminus\Lambda_k}$ signifies a trace over the Hilbert spaces corresponding to times $t_\alpha\in\Lambda_\ell\setminus\Lambda_k$ and $\mathbb{1}_{\Lambda_k}$ is the identity matrix on the remaining Hilbert space $\mathcal{H}_{\Lambda_k} := \bigotimes_{\alpha\in\Lambda_k}(\mathcal{H}^{(2)}_\alpha\otimes\mathcal{H}^{(1)}_\alpha)$.

A classical family of joint probability distributions $\mathbb{P}_{\Lambda_k}$ can be represented by a family of classical combs [56]

$$\Upsilon^{\text{cl}}_{\Lambda_k} = \sum_{i_k,\dots,i_1}\mathbb{P}_{\Lambda_k}(i_k,t_k;\dots;i_1,t_1)\mathbb{1}^{(2)}_{t_k}\otimes|i_k\rangle\langle i_k|\otimes\cdots\otimes\mathbb{1}^{(2)}_{t_1}\otimes|i_1\rangle\langle i_1|, \tag{B3}$$

where $\mathbb{1}^{(2)}_{t_\alpha}\in\mathcal{B}(\mathcal{H}^{(2)}_\alpha)$ are identity matrices and $|i_\alpha\rangle\langle i_\alpha|\in\mathcal{B}(\mathcal{H}^{(1)}_\alpha)$ are orthogonal pure states corresponding to the measurement outcomes i_α. The classical combs defined by (B3) are diagonal in the reference basis and correctly reproduce the probabilities given by $\mathbb{P}_{\Lambda_k}$; indeed, it is easy to see that we have

$$\text{tr}\left[\left(P^{\text{T}}_{i_k}\otimes\cdots\otimes P^{\text{T}}_{i_1}\right)\Upsilon^{\text{cl}}_{\Lambda_k}\right] = \mathbb{P}(i_k,t_k;\dots;i_1,t_1). \tag{B4}$$

As $\mathrm{tr}\left[\left(\sum_{j_\alpha} P_{j_\alpha}\right) P_{i_\alpha}\right] = \mathrm{tr}\left[\Phi_\alpha^+ P_{i_\alpha}\right]$, the consistency property of the joint probability distributions induces a consistency property of the family of classical combs $\left\{\Upsilon^{\mathrm{cl}}_{\Lambda_k}\right\}$ constructed via to Eq. (B3). Then, according to the GET, there exists a classical comb $\Upsilon^{\mathrm{cl}}_{\Lambda}$ that has all the finite combs $\Upsilon^{\mathrm{cl}}_{\Lambda_k}$ as 'marginals', *i.e.*, $\Upsilon^{\mathrm{cl}}_{\Lambda_k} = \Upsilon^{\mathrm{cl}|\Lambda_k}_{\Lambda}$ for all finite $\Lambda_k \subseteq \Lambda$. This implies the existence of a joint probability distribution $\mathbb{P}_\Lambda$ that has all finite $\mathbb{P}_{\Lambda_k}$ as marginals, which proves the original KET. □

---

[1] Z. Schuss, *Theory and Applications of Stochastic Processes: An Analytical Approach* (Springer, New York, 2009).
[2] M. Liao, *Applied Stochastic Processes* (Chapman and Hall/CRC, Boca Raton, 2013).
[3] A. N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung* (Springer, Berlin, 1933) [*Foundations of the Theory of Probability* (Chelsea, New York, 1956)].
[4] W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1971).
[5] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2007).
[6] T. Tao, *An Introduction to Measure Theory* (American Mathematical Society, 2011).
[7] A. J. Leggett and A. Garg, Phys. Rev. Lett. **54**, 857 (1985).
[8] A. J. Leggett, Rep. Prog. Phys. **71**, 022001 (2008).
[9] C. Emary, N. Lambert, and F. Nori, Rep. Prog. Phys. **77**, 016001 (2014).
[10] N. Gilbert, *Agent-Based Models* (SAGE Publications Inc., Los Angeles, 2007).
[11] J. Pearl, *Causality: Models, Reasoning and Inference* (Cambridge University Press, Cambridge, U.K.; New York, 2009).
[12] S. Milz, F. A. Pollock, and K. Modi, Open Sys. Info. Dyn. , 1740016 (2017).
[13] F. Costa and S. Shrapnel, New J. Phys. **18**, 063032 (2016).
[14] O. Oreshkov and C. Giarmatzi, New J. Phys. **18**, 093020 (2016).
[15] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Phys. Rev. X **7**, 031021 (2017).
[16] H.-P. Breuer, E.-M. Laine, J. Piilo, and B. Vacchini, Rev. Mod. Phys. **88**, 021002 (2016).
[17] A. Smirne, D. Egloff, M. G. Díaz, M. B. Plenio, and S. F. Huelga, arXiv:1709.05267 (2017).
[18] L. Accardi, A. Frigerio, and J. T. Lewis, Publ. Rest. Inst. Math. Sci. **18**, 97 (1982).
[19] N. Wiener, A. Siegel, B. Rankin, and W. T. Martin, eds., *Differential Space, Quantum Systems, and Prediction* (The MIT Press, Cambridge (MA), 1966).
[20] M. P. Lévy, Am. J. Math. **62**, 487 (1940).
[21] Z. Ciesielski, *Lectures on Brownian motion, heat conduction and potential theory* (Aarhus Universitet, Mathematisk Institutt, 1966).
[22] R. Bhattacharya and E. C. Waymire, *A Basic Course in Probability Theory* (Springer, New York, NY, 2017).
[23] S. Shrapnel and F. Costa, Quantum **2**, 63 (2018).
[24] G. Lindblad, Comm. Math. Phys. **65**, 281 (1979).
[25] E. B. Davis, *Quantum Theory of Open Systems* (Academic Press Inc, London; New York, 1976).
[26] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).
[27] S. Shrapnel, F. Costa, and G. Milburn, New J. Phys. **20**, 053010 (2018).
[28] K. Modi, Sci. Rep. **2**, 581 (2012).
[29] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Phys. Rev. A **97**, 012127 (2018).
[30] S. Milz, F. A. Pollock, and K. Modi, arXiv:1610.02152 (2016).
[31] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Europhys. Lett. **83**, 30004 (2008).
[32] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. Lett. **101**, 060401 (2008).
[33] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).
[34] C. M. Caves, C. A. Fuchs, K. K. Manne, and J. M. Renes, Found. Phys. **34**, 193 (2004).
[35] O. Oreshkov, F. Costa, and Č. Brukner, Nat. Commun. **3**, 1092 (2012).
[36] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and Č. Brukner, New J. Phys. **17**, 102001 (2015).
[37] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, Phys. Rev. A **88**, 022318 (2013).
[38] R. W. Spekkens, Phys. Rev. A **71**, 052108 (2005).
[39] E. G. Cavalcanti, Phys.Rev. X **8**, 021018 (2018).
[40] D. Kretschmann and R. F. Werner, Phys. Rev. A **72**, 062323 (2005).
[41] F. Caruso, V. Giovannetti, C. Lupo, and S. Mancini, Rev. Mod. Phys. **86**, 1203 (2014).
[42] C. Portmann, C. Matt, U. Maurer, R. Renner, and B. Tackmann, IEEE Trans. Inf. Theory **63**, 3277 (2017).
[43] L. Hardy, arXiv:1608.06940 (2016).
[44] L. Hardy, Phil. Trans. R. Soc. A **370**, 3385 (2012).
[45] J. Cotler, C.-M. Jian, X.-L. Qi, and F. Wilczek, arXiv:1711.03119 (2017).
[46] N. Barnett and J. P. Crutchfield, J. Stat. Phys. **161**, 404 (2015).
[47] J. Thompson, A. J. P. Garner, V. Vedral, and M. Gu, npj Quantum Inf. **3**, 6 (2017).
[48] F. A. Pollock, C. Rodríguez-Rosario, T. Frauenheim, M. Paternostro, and K. Modi, Phys. Rev. Lett. **120**, 040405 (2018).
[49] R. Tumulka, Lett. Math. Phys. **84**, 41 (2008).
[50] E. Haapasalo, T. Heinosaari, and Y. Kuramochi, J. Phys. A **49**, 33LT01 (2016).

[51] S. Milz, F. A. Pollock, T. P. Le, G. Chiribella, and K. Modi, New J. Phys. **20**, 033033 (2018).
[52] J. v. Neumann, Comp. Math. **6**, 1 (1939).
[53] M. Reed and B. Simon, *Functional Analysis* (Academic Press, New York, 1980).
[54] M.-D. Choi, Linear Algebra Appl. **10**, 285 (1975).
[55] A. Jamiołkowski, Rep. Math. Phys. **3**, 275 (1972).
[56] We will generally drop the explicit distinction between the comb $\mathcal{T}_{\Lambda_k}$ and its Choi state Υ_{Λ_k}.