

Information and disturbance in operational probabilistic theories

Giacomo Mauro D’Ariano,^{1,*} Paolo Perinotti,^{1,†} and Alessandro Tosini^{1,‡}

¹*QUIT group, Physics Dept., Pavia University, and INFN Sezione di Pavia, via Bassi 6, 27100 Pavia, Italy*

Any measurement is intended to provide *information* on a system, namely knowledge about its state. However, we learn from quantum theory that it is generally impossible to extract information without disturbing the state of the system or its correlations with other systems. In this paper we address the issue of the interplay between information and disturbance for a general operational probabilistic theory. The traditional notion of disturbance considers the fate of the system state after the measurement. However, the fact that the system state is left untouched ensures that also correlations are preserved only in the presence of local discriminability. Here we provide the definition of disturbance that is appropriate for a general theory. We then prove an equivalent condition for no-information without disturbance—*atomicity of the identity*—namely the impossibility of achieving the trivial evolution—the *identity*—as the coarse-graining of a set of non trivial ones. We prove a general theorem showing that information that can be retrieved without disturbance corresponds to perfectly repeatable and discriminating tests. As a consequence we prove a structure theorem for operational probabilistic theories, showing that the set of states of any system decomposes as a direct sum of perfectly discriminable sets, and such decomposition is preserved under system composition. Besides proving that no-information without disturbance is implied by the purification postulate, we show via concrete examples that the converse is not true. Finally we show that no-information without disturbance and local discriminability are independent.

I. INTRODUCTION

The possibility that gathering information on a physical system may affect the state of the system itself was introduced by Heisenberg in his famous gedanken experiment [1], which became the first paradigm of quantum mechanics. The issue raised by Heisenberg spawned a vast literature up to present days (see [2, 3] as recent reviews), with a variety of quantifications of “information” and “disturbance” and corresponding tradeoff relations [4–7]. All these results are quantitative accounts of a core issue in quantum theory, the *no-information without disturbance* theorem [8, 9]. The proofs of the theorem rely on the mathematical structure of quantum theory, and thus do not emphasise the logical relation between no-information without disturbance and other quantum features, such as *local discriminability* (the possibility of discriminating multipartite states via only local measurements) or *purification* (every mixed state can be obtained as the marginal state of a pure state).

The framework here used for exploring the relation between information and disturbance is that of *operational probabilistic theories* (OPTs) [9–11]. In this setting a rigorous formulation of the notions of system, process, and their compositions is given, which constitutes the grammar for the probabilistic description of an experiment. Quantum theory and classical theory are two instances of OPTs.

For some probabilistic theories which can be reframed as OPTs, the definitions of information and disturbance

have been investigated in the presence of local discriminability, purification, and causality [12–15]. For OPTs satisfying those three axioms the no-information without disturbance theorem has been proved in Refs. [9, 10]. In the present paper we point out a weakness in the existing notion of disturbance, which is ubiquitous in all past approaches. Indeed, the conventional definition of disturbance asserts that an experiment does not disturb the system if and only if its overall effect is to leave its state unchanged, disregarding the effects of the experiment on the environment. Whilst this captures the meaning of disturbance within quantum theory, we cannot consistently apply the same notion in theories that violate local discriminability. A significative case is that of the Fermionic theory [16–18] where, due to the parity superselection rule, an operation that does not disturb a bunch of Fermionic systems still could affect their correlations with other systems. This issue can be cured asking a non-disturbing experiment to preserve not only the system state, but also its purifications [9, 10]. This extension of the notion of disturbance is general enough to capture the operational meaning of disturbance for Fermionic systems, however, it is still unsatisfactory, since it cannot be used to describe disturbance in models that do not enjoy purification, e. g. classical information theory.

Here we will define non-disturbing operations only by referring to the OPT framework, thus providing a notion that holds also for theories that do not satisfy local discriminability, purification, or causality, and even for theories whose sets of states are not convex. Given a system, and an operation on it, the fate of any possible dilation of the states of the system is taken into account, where by *dilation* we mean any state of a larger system, whose marginal is the dilated state [19]. We prove then a necessary and sufficient condition for a theory to satisfy no-information without disturbance. The condition is the

* dariano@unipv.it

† paolo.perinotti@unipv.it

‡ alessandro.tosini@unipv.it

impossibility of realizing the identity transformation as a nontrivial coarse-graining of a set of operations. Technically speaking the above conditions amounts to *atomicity* of the identity. Moreover, since a theory might satisfy no-information without disturbance only when restricted to some collections of states, we will provide a weaker necessary and sufficient condition for this case.

Similarly to the Heisenberg uncertainty relations, the no-information without disturbance has been considered as a characteristic quantum trait. Instead, as we will see here, this feature can be exhibited in the absence of most of the principles of quantum theory [9], and it is ubiquitous among OPTs. Moreover, the most general case is that of an OPT where *some* information can be extracted without disturbance, in which case this information has all the features of a classical one. On the other hand, the only kind of systems that allow extracting any information without disturbance is the classical system. This observation provides an alternative way of characterising classical systems with respect to Ref. [20].

In Section II we review the framework of operational probabilistic theories and some relevant features that characterize quantum theory within this scenario. In Section III we generalize the notion of *equality upon-input* to general OPTs, including the cases in which local discriminability does not hold. In Section IV, after introducing the definition of information and disturbance, we present the main results of this paper: i) the atomicity of the identity evolution as a necessary and sufficient condition for no-information without disturbance; ii) other equivalent necessary and sufficient conditions in terms of properties of reversible evolutions of the theory; iii) other only sufficient conditions (e.g. purification); iv) a structure theorem for theories where some information can be extracted without disturbance. We prove that the information that can be extracted without disturbance is “classical”, in the sense that the measurement is a repeatable reading of shareable information. Moreover, the classical theory of information is the only OPT with local discriminability in which all the information can be extracted without disturbance. In Section V we deepen the relation between no-information without disturbance and other characteristic properties of quantum theory. We show that no-information without disturbance can be satisfied by theories without purification and independently of local discriminability, providing counterexamples based on some of the conditions mentioned above. We end with the conclusions in Section VI.

II. THE FRAMEWORK

In this section we review the framework of *operational probabilistic theories* (OPT) (we refer to [9–11] for further details).

The primitives of an operational theory are the notions of *test*, *event*, and *system*. A test $\{\mathcal{A}_i\}_{i \in X}$ is the collection of events \mathcal{A}_i , where i labels the element of the

outcome space X . In the quantum case \mathcal{A}_i is the i th quantum operation of the quantum instrument $\{\mathcal{A}_i\}_{i \in X}$. The notion of test bridges the experiment with the theory, with $i \in X$ denoting the objective outcome, and \mathcal{A}_i the mathematical description of the corresponding event. The notion of *system*, here denoted by capital Roman letters A, B, \dots , rules connections of tests. An input and an output label are associated to any test (event). We represent a test $\mathcal{A}_X := \{\mathcal{A}_i\}_{i \in X}$ and its building events \mathcal{A}_i by the diagrams

$$\begin{array}{c} \text{---} A \quad \boxed{\mathcal{A}_X} \quad \text{---} B \\ \text{---} A \quad \boxed{\mathcal{A}_i} \quad \text{---} B \end{array},$$

respectively, with the rule that an output wire can be connected only to an input wire with the same label. Thus, given two tests \mathcal{A}_X and \mathcal{B}_Y we can define their *sequential composition* $(\mathcal{B}\mathcal{A})_{X \times Y}$ as the collection of events

$$\begin{array}{c} \text{---} A \quad \boxed{\mathcal{B}_j \mathcal{A}_i} \quad \text{---} C \\ \text{---} A \quad \boxed{\mathcal{A}_i} \quad \text{---} B \quad \boxed{\mathcal{B}_j} \quad \text{---} C \end{array},$$

for $i \in X$ and $j \in Y$. A *singleton test* is a test containing a single event. We call such an event *deterministic*. For every system A there exists a unique singleton test $\{\mathcal{I}_A\}$ such that $\mathcal{I}_B \mathcal{A} = \mathcal{A} \mathcal{I}_A = \mathcal{A}$ for every event \mathcal{A} with input A and output B , and we call \mathcal{I}_A *identity* of system A . Besides sequential compositions of tests and events, a theory is specified by the rule for composing them in parallel. For every couple of systems (A, B) we can form the composite system $C := AB$, on which we can perform tests $(\mathcal{C} \otimes \mathcal{D})_{X \times Y}$ with events $\mathcal{C}_i \otimes \mathcal{D}_j$ in *parallel composition* represented as follows

$$\begin{array}{c} \text{---} A \quad \boxed{\mathcal{C}_i \otimes \mathcal{D}_j} \quad \text{---} B \\ \text{---} C \quad \quad \quad \text{---} D \end{array} = \begin{array}{c} \text{---} A \quad \boxed{\mathcal{C}_i} \quad \text{---} B \\ \text{---} C \quad \boxed{\mathcal{D}_j} \quad \text{---} D \end{array},$$

and satisfying the condition $(\mathcal{E}_h \otimes \mathcal{F}_k)(\mathcal{C}_i \otimes \mathcal{D}_j) = (\mathcal{E}_h \mathcal{C}_i) \otimes (\mathcal{F}_k \mathcal{D}_j)$. Notice that we use the tensor product symbol \otimes for the parallel composition rule. Actually, for the quantum and the classical OPT the parallel composition is the usual tensor product of linear maps. However, for a general OPT, the parallel composition may not coincide with a tensor product.

There exists a special system type I, the *trivial system*, such that $AI = IA = A$ for every system A . The tests with input system I and output A are called *preparation-tests* of A , while the tests with input system A and output I are called *observation-tests* of A . Preparation-events of A are graphically denoted as boxes without the input wire $\overline{\rho} \boxed{A}$ (or in formula as round kets $|\rho\rangle_A$), and the observation-events by boxes with no output wire $\overline{A} \boxed{C}$ (in formula round bras $\langle c|_A$). For example, one can have events of the following kind

$$\begin{array}{c} \overline{\rho_i} \boxed{B} \\ \text{---} C \quad \boxed{\mathcal{D}_j} \quad \text{---} D \end{array} = \text{---} C \quad \boxed{\rho_i \otimes \mathcal{D}_j} \quad \text{---} BD.$$

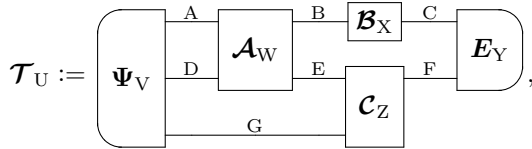
We will always use the Greek letters to denote preparation-tests $\{\rho_i\}_{i \in X}$ and Latin letters to denote

observation-tests $\{c_j\}_{j \in X}$ (we will not specify the system when it is clear from the context).

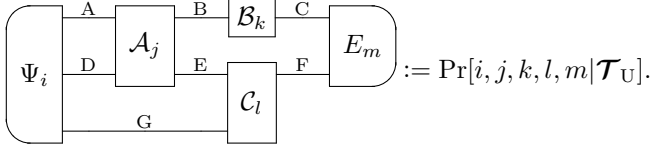
An arbitrary test obtained by parallel and sequential composition of box diagrams is called *circuit*. A circuit is *closed* if its overall input and output systems are trivial: it starts with a preparation test and ends with an observation test. An operational theory is *probabilistic* if to any closed circuit of tests corresponds a probability distribution for the joint test. Such a theory is an OPT. For example the application of an observation-event c_i after the preparation-event ρ_j corresponds to the closed circuit $(c_i | \rho_j)_A$ and denotes the probability of the outcome (i, j) of the observation-test c_X after the preparation-test ρ_Y of system A, i.e.

$$(\rho_j | \rho_j)_A (c_i) := \Pr [i, j | (\rho_Y)_A (c_X)].$$

For a more complex example, consider the test



with $U = V \times W \times X \times Y \times Z$. Then we define



Summarising: by a closed circuit made of events we denote their joint probability upon the connection specified by the circuit graph, with nodes being the test boxes, and links being the system wires.

Given a system A of a probabilistic theory we can quotient the set of preparation-events of A by the equivalence relation $|\rho\rangle_A \sim |\sigma\rangle_A \Leftrightarrow (c|\rho)_A = (c|\sigma)_A$ for every observation-event c . Similarly we can quotient observation-events. The equivalence classes of preparation-events and observation-events of A will be denoted by the same symbols as their elements $|\rho\rangle_A$ and $(c|_A$, respectively, and will be called *state* and *effect* for system A.

For every system A, we will denote by $\text{St}(A)$, $\text{Eff}(A)$ the sets of states and effects, respectively. States and effects are real-valued functionals on each other, and can be naturally embedded in reciprocally dual real vector spaces, $\text{St}_{\mathbb{R}}(A)$ and $\text{Eff}_{\mathbb{R}}(A)$, whose dimension $\dim(A)$ is assumed to be finite.

In Appendix A it is proved that an event \mathcal{A} with input system A and output system B induces a linear map from $\text{St}_{\mathbb{R}}(AC)$ to $\text{St}_{\mathbb{R}}(BC)$ for each ancillary system C. The collection of all these maps is called *transformation* from A to B. More explicitly, given two transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$, one has $\mathcal{A} = \mathcal{A}'$, if and only if

$$\left(\Psi \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array} \begin{array}{c} \mathcal{A} \\ \mathcal{A}' \end{array} \right) a = \left(\Psi \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array} \begin{array}{c} \mathcal{A}' \\ \mathcal{A} \end{array} \right) a,$$

for every C, every $\Psi \in \text{St}(AC)$, and every $a \in \text{Eff}(BC)$, namely they give the same probabilities within every possible closed circuit. Notice that, using the fact that two states are equal if and only if they give the same probability when paired to every effect, the above condition amounts to state that $\mathcal{A} = \mathcal{A}'$ if and only if

$$\left(\Psi \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array} \begin{array}{c} \mathcal{A} \\ \mathcal{A}' \end{array} \right) = \left(\Psi \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array} \begin{array}{c} \mathcal{A}' \\ \mathcal{A} \end{array} \right), \quad (1)$$

for every C, and every $\Psi \in \text{St}(AC)$.

In the following, the symbols \mathcal{A} and $\begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array} \mathcal{A} \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \end{array}$ will be used to represent the transformation corresponding to the event \mathcal{A} . The set of transformations from A to B will be denoted by $\text{Transf}(A, B)$, with linear span $\text{Transf}_{\mathbb{R}}(A, B)$. It is now obvious that a linear map $\mathcal{A} \in \text{Transf}_{\mathbb{R}}(A, B)$ is *admissible* if it locally preserves the set of states $\text{St}(AC)$, namely $\mathcal{A} \otimes \mathcal{I}_C(\text{St}(AC)) \subseteq \text{St}(BC)$, for every system C. In the following we will write $\mathcal{A} | \Psi_{AC}$ instead of $\mathcal{A} \otimes \mathcal{I}_C | \Psi_{AC}$, with $\Psi \in \text{St}(AC)$ and $\mathcal{A} \in \text{Transf}(A, B)$ when the domains are clear from the context.

An operational probabilistic theory is now defined as a collection of systems and transformations with the above rules for parallel and sequential composition and with a probability associated to any closed circuit[21].

We introduce now the notions of *refinement* of an event and *atomic* event.

Definition 1 (Refinement of an event). *A refinement of an event $\mathcal{C} \in \text{Transf}(A, B)$ is given by a collection of events $\{\mathcal{D}_i\}_{i \in X}$ from A to B, such that there exists a test $\{\mathcal{D}_i\}_{i \in Y}$ with $X \subseteq Y$ and $\mathcal{C} = \sum_{i \in X} \mathcal{D}_i$. We say that a refinement $\{\mathcal{D}_i\}_{i \in X}$ of \mathcal{C} is trivial if $\mathcal{D}_i = \lambda_i \mathcal{C}$, $\lambda_i \in [0, 1]$, for every $i \in X$. Conversely, \mathcal{C} is called the coarse-graining of the events $\{\mathcal{D}_i\}_{i \in X}$.*

In the following we will often refer to a refinement of \mathcal{C} simply as $\mathcal{C} = \sum_{i \in X} \mathcal{D}_i$, without specifying the test including the events \mathcal{D}_i .

Definition 2 (Refining event). *Given two events $\mathcal{C}, \mathcal{D} \in \text{Transf}(A, B)$ we say that \mathcal{D} refines \mathcal{C} , and write $\mathcal{D} \prec \mathcal{C}$, if there exist a refinement $\{\mathcal{D}_i\}_{i \in X}$ of \mathcal{C} such that $\mathcal{D} \in \{\mathcal{D}_i\}_{i \in X}$.*

Definition 3 (Non redundant test). *We call a test $\{\mathcal{A}_i\}_{i \in X}$ non redundant when for every pair $i, j \in X$ one has $\mathcal{A}_i \neq \lambda \mathcal{A}_j$ for $\lambda > 0$.*

Notice that a test that is redundant can be interpreted as a non redundant test followed by a conditional coin tossing. As a consequence a redundant test always gives some spurious information, unrelated to the input state. From a redundant test one can achieve a maximal non redundant one by taking the test made of coarse grainings of all the sets of proportional elements.

Definition 4 (Refinement set). *Given an event $\mathcal{C} \in \text{Transf}(A, B)$ we define its refinement set $\text{Ref}_{\mathcal{C}}$ the set of all events that refine \mathcal{C} .*

Definition 5 (Atomic and refinable events). *An event \mathcal{C} is atomic if it admits only trivial refinements, namely $\mathcal{D} \prec \mathcal{C}$ implies $\mathcal{D} = \lambda\mathcal{C}$, $\lambda \in [0, 1]$. An event is refinable if it is not atomic.*

In the special case of states, the word *pure* is used as synonym of atomic, with a pure state describing an event that provides maximal knowledge about the system's preparation. This means that the knowledge provided by a pure state cannot be further refined. As usual a state that is not pure will be called *mixed*.

A relevant notion is that of *internal* state:

Definition 6 (Internal state). *Given a system A , a state $\chi \in \text{St}(A)$ is internal if for every state $\rho \in \text{St}(A)$ there is a non-zero probability $p > 0$ such that $p\rho \in \text{Ref}_\chi$.*

The geometric interpretation of “internal” is the intuitive one, namely an internal state cannot belong to the boundary of the set of states $\text{St}(A)$.

The last definition we introduce is that of state dilation.

Definition 7 (Dilation of a state ρ). *We say that $\Psi \in \text{St}(AB)$ is a dilation of $\rho \in \text{St}(A)$ if*

$$\rho \text{---} \text{A} = \left(\Psi \begin{array}{l} \text{A} \\ \text{B} \end{array} \begin{array}{l} \text{---} \\ \text{---} \end{array} \begin{array}{l} \text{---} \\ e_B \end{array} \right)$$

for some deterministic effect $e \in \text{Eff}(B)$. We denote by D_ρ the set of all dilations of the state ρ . More generally, given a collection of states $X \subseteq \text{St}(A)$ we define $D_X := \bigcup_{\rho \in X} D_\rho$.

We remark that, given $\sigma \in X$, every state of the form $\sigma \otimes \rho$ belongs to D_X .

Notice that there are generally more than one deterministic effect for the same system, differently from quantum theory, where the partial trace over the Hilbert space of the system is the only way to discard it. Instead, in a theory with more deterministic effects for the same system B the marginal state of system A generally depends on the effect used to discard the system B . In the following we will call *marginal of a state with deterministic effect e* the specific marginal obtained by applying the effect $e \in \text{Eff}(B)$.

Lemma 1. *Given a state $\rho \in \text{St}(A)$ one has $\text{Ref}_{D_\rho} \subseteq D_{\text{Ref}_\rho}$, where Ref_{D_ρ} denotes the union of the refinements of any state in D_ρ .*

Proof. Consider a system B and a state $\Psi \in \text{St}(AB)$:

1. $\Psi \in D_{\text{Ref}_\rho}$ iff $(e|_B|\Psi)_{AB} = |\sigma\rangle_A$, with $\sigma \in \text{Ref}_\rho$, and $e \in \text{Eff}(B)$ deterministic.
2. $\Psi \in \text{Ref}_{D_\rho}$ iff there exists a state $\Omega \in \text{St}(AB)$, with $\Omega \in D_\rho$, and a refinement $\{\Psi_i\}_{i \in X}$ of Ω such that $\Psi \in \{\Psi_i\}_{i \in X}$.

By hypothesis $\{\Psi_i\}_{i \in X}$ is a refinement of Ω , namely

$$|\Omega\rangle_{AB} = \sum_{i \in X} |\Psi_i\rangle_{AB}, \quad (2)$$

and $\Omega \in D_\rho$, namely $(e|_B|\Omega)_{AB} = |\rho\rangle_A$. Accordingly, marginalising both sides of Eq. (2) one has that for every $i \in X$, $(e|_B|\Psi_i)_{AB} = |\sigma_i\rangle_A \in \text{Ref}_\rho$. Since $\Psi \in \{\Psi_i\}_{i \in X}$, this concludes the proof. ■

A. Relevant classes of OPTs

A frequently highlighted property within the wider scenario of OPTs is that of multipartite states discrimination via local measurements:

Definition 8 (Local discriminability). *It is possible to discriminate between any pair of states of composite systems using only local measurements. Mathematically, given two joint states $\Psi, \Psi' \in \text{St}(AB)$ with $\Psi \neq \Psi'$, there exist two effects $a \in \text{Eff}(A)$ and $b \in \text{Eff}(B)$, such that*

$$\left(\Psi \begin{array}{l} \text{A} \\ \text{B} \end{array} \begin{array}{l} a \\ b \end{array} \right) \neq \left(\Psi' \begin{array}{l} \text{A} \\ \text{B} \end{array} \begin{array}{l} a \\ b \end{array} \right).$$

Two relevant consequences of local discriminability are: i) the local characterization of transformations, stating that the local behaviour of a transformation is sufficient to fully characterize the transformation itself; ii) the atomicity of parallel composition. Here we report those two features for the convenience of the reader.

Proposition 1 (Local characterization of transformations). *If local discriminability holds, then for any two transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$, the condition $\mathcal{A}|\rho\rangle_A = \mathcal{A}'|\rho\rangle_A$ for every $\rho \in \text{St}(A)$ implies $\mathcal{A} = \mathcal{A}'$.*

See Ref. [10] for the proof.

Proposition 2 (Atomicity of parallel composition). *If an OPT satisfies local discriminability then the parallel composition is atomic.*

For the proof of the above proposition see Ref. [22]. We observe that an OPT with local discriminability allows for tomography of multipartite states using only local measurements. In an OPT with local discriminability, the linear space of effects of a composite system is the tensor product of the linear spaces of effects of the component systems, namely $\text{Eff}(AB)_{\mathbb{R}} \equiv \text{Eff}(A)_{\mathbb{R}} \otimes \text{Eff}(B)_{\mathbb{R}}$. Thus, any bipartite effect $c \in \text{Eff}(AB)$ can be written as a linear combination of product effects, and every probability $(c|\rho)_{AB}$, for $\rho \in \text{St}(AB)$, can be computed as a linear combination of the probabilities $((a|_A \otimes (b|_B)|\rho)_{AB}$ arising from a finite set of product effects. The same holds for the linear space of states and in an OPT with local discriminability the parallel composition of two states (effects) can be understood as a tensor product. Finally, the relation $\dim(AB) = \dim(A)\dim(B)$ between

the linear dimension of the set of states/effects holds, whereas for theories without local discriminability it holds $\dim(AB) > \dim(A)\dim(B)$.

Recently it has been shown that relevant physical theories, such as the Fermionic theory [16], can be described in the OPT framework relaxing the property of local discriminability [17, 18]. The most general scenario for OPTs that exhibit a finite degree of holism is that of OPTs with n -local discriminability for some $n \in \mathbb{N}$ [23]:

Definition 9 (*n*-local discriminability). *A theory satisfies n-local discriminability if whenever two states ρ and ρ' are different, there exist a n-local effect b such that $(b|\rho) \neq (b|\rho')$. We say that an effect is n-local if it can be written as a conic combination of tensor products of effects that are at most n-partite.*

Two notable examples are indeed Fermionic quantum theory and real quantum computation [17, 18, 23] that are both 2-local tomographic.

Another relevant class of OPTs is that of theories with purification [10, 24]. As a result of this paper we will show (Proposition 7) that the set of OPTs with purification is strictly smaller than the set of OPTs that satisfy no-information without disturbance. Moreover, we will see that a weak version of purification, which does not require the uniqueness (as in quantum theory) but just the existence of a purification for each state of the theory, is enough to imply no-information without disturbance. Accordingly, we define the following class of OPTs.

Definition 10 (Purification). *We say that an OPT satisfies purification if for every system A and for every state $\rho \in \text{St}(A)$, there exists a system B and a pure state $\Psi \in \text{St}(AB)$ that is a dilation of ρ .*

Remark 1 (Equivalent definition of purification). One can provide an equivalent definition of purification based on the notion of dilation of Definition 7. Indeed, an OPT satisfies purification if for every system A and for every state $\rho \in \text{St}(A)$, the dilation set D_ρ contains a pure state.

As already noticed, the above definition does not require the purification of a state to be unique up to reversible transformations on the purifying system, as it is for example in quantum theory.

The last relevant class of OPTs that we point out is that of causal theories:

Definition 11 (Causal OPTs). *The probability of preparation events in a closed circuit is independent of the choice of observations.*

Mathematically, if $\{\rho_i\}_{i \in X} \subset \text{St}(A)$ is a preparation-test, then the conditional probability of the preparation ρ_i given the choice of the observation-test $\{a_j\}_{j \in Y}$ is the marginal

$$\Pr(i|\{a_j\}) := \sum_{j \in Y} (a_j|\rho_i)_A.$$

In a causal theory the marginal probability $\Pr(i|\{a_j\})$ is independent of the choice of the observation-test $\{a_j\}$: if $\{a_j\}_{j \in Y}$ and $\{b_k\}_{k \in Z}$ are two different observation-tests, then one has $\Pr(i|\{a_j\}) = \Pr(i|\{b_k\})$.

The present notion of causality is simply the *Einstein causality* expressed in the language of OPTs. As proved in Ref. [10] causality is equivalent to the existence of a unique deterministic effect e_A . We call the effect e_A the *deterministic effect* for system A. By definition in non causal theories the deterministic effect cannot be unique.

III. OPERATIONAL IDENTITIES BETWEEN TRANSFORMATIONS

As expressed in Eq. (1), two transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$ of an OPT are said to be operationally identical if for every system C and for every state $\Psi \in \text{St}(AC)$ one has $\mathcal{A}|\Psi)_{AC} = \mathcal{A}'|\Psi)_{AC}$. However, two non-identical maps $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$ could behave in the same way when their action is restricted to a relevant subclass of states.

The notion of identical transformation upon input of a state $\rho \in \text{St}(A)$ has been introduced in the literature (see Refs. [9, 25] and references therein):

Definition 12 (Identical transformations upon input of ρ). *We say that two transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$ are equal upon input of $\rho \in \text{St}(A)$, and write $\mathcal{A} =_\rho \mathcal{A}'$, if for every $\sigma \in \text{Ref}_\rho$ it is $\mathcal{A}\sigma = \mathcal{A}'\sigma$.*

According to this definition, even if $\mathcal{A} =_\rho \mathcal{A}'$, still the maps \mathcal{A} and \mathcal{A}' could act differently on dilations of ρ , namely it could be $\mathcal{A}|\Psi)_{AC} \neq \mathcal{A}'|\Psi)_{AC}$, for some $\Psi \in \text{St}(AC)$ with $\Psi \in D_\rho$. In this case the difference between \mathcal{A} and \mathcal{A}' would go undetected if only their action on system A is considered. As proved in Corollary 1, the local action of a map is sufficient to determine the map itself if the OPT satisfies local discriminability. However, for theories without local discriminability the local action of a transformation might not be sufficient to characterize it. For this reason we introduce the notion of identical transformation upon input of dilations of a state ρ .

Definition 13 (Identical transformations upon input of D_ρ). *Given a state $\rho \in \text{St}(A)$, we say that two transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$ are equal upon input of D_ρ , and write $\mathcal{A} =_{D_\rho} \mathcal{A}'$, if $\mathcal{A}|\Psi)_{AC} = \mathcal{A}'|\Psi)_{AC}$ for every $\Psi \in D_{\text{Ref}_\rho}$.*

An immediate consequence of the above definition is the following lemma

Lemma 2 (Identical transformations). *If two transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$ are equal upon input of D_χ , with $\chi \in \text{St}(A)$ internal, then the two transformations are identical and we simply write $\mathcal{A} = \mathcal{A}'$.*

Proof. This immediately follows observing that if χ is an internal state of system A then D_χ coincides with the set of all states $\text{St}(AC)$ for any possible system C. ■

In Definition 5 we introduced the notion of atomic events. Based on the present notion of identical transformation upon input of D_ρ we can provide a weaker version of atomicity for transformations

Definition 14 (Atomic and refinable transformation upon input of D_ρ). *A transformation $\mathcal{A} \in \text{Transf}(A, B)$ is atomic upon input of D_ρ , with $\rho \in \text{St}(A)$, if all its refinements are trivial upon input of D_ρ , namely $\mathcal{B} \prec \mathcal{A}$ implies $\mathcal{B} =_{D_\rho} \lambda \mathcal{A}$, $\lambda \in [0, 1]$. Conversely, we say that an event is refinable upon input of D_ρ whenever it is not atomic upon input of D_ρ .*

Here we show that the two definitions 12 and 13 coincide for causal OPTs with local discriminability. For this purpose we first need the following lemma.

Lemma 3. *In a causal OPT, if $\Psi \in \text{St}(AB)$, with $\Psi \in D_{\text{Ref}_\rho}$ for some $\rho \in \text{St}(A)$, then*

$$\{(b|_B | \Psi)_{AB} | b \in \text{Eff}(B)\} \subseteq \text{Ref}_\rho. \quad (3)$$

Proof. By definition it is $(e|_B | \Psi)_{AB} = |\sigma\rangle_A$ with $\sigma \in \text{Ref}_\rho$, with e the unique deterministic effect of B . Given an arbitrary effect $b \in \text{Eff}(B)$, we can always consider the observation test $\{b, e - b\}$. Since it is $(b|_B | \Psi)_{AB}, (e - b|_B | \Psi)_{AB} \in \text{Ref}_\rho$ we have that $\{(b|_B | \Psi)_{AB} | b \in \text{Eff}(B)\} \subseteq \text{Ref}_\rho$. ■

Proposition 3. *In a causal OPT with local discriminability, given two transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, B)$, the two conditions $\mathcal{A} =_\rho \mathcal{A}'$ and $\mathcal{A} =_{D_\rho} \mathcal{A}'$ are equivalent.*

Proof. We first prove that $\mathcal{A} =_\rho \mathcal{A}' \Rightarrow \mathcal{A} =_{D_\rho} \mathcal{A}'$. Consider an arbitrary $\Psi \in D_{\text{Ref}_\rho}$, with $\rho \in \text{St}(A)$. Let for example be $\Psi \in \text{St}(AC)$. By hypothesis it is $\mathcal{A} =_\rho \mathcal{A}'$, namely $\mathcal{A}|\sigma\rangle_A = \mathcal{A}'|\sigma\rangle_A$ for every $\sigma \in \text{Ref}_\rho$. Then, due to Lemma 3, $\forall b \in \text{Eff}(B), \forall c \in \text{Eff}(C)$, it is $(b|_B (c|_C (\mathcal{A} \otimes \mathcal{I}_C) | \Psi)_{AC} = (b|_B (c|_C (\mathcal{A}' \otimes \mathcal{I}_C) | \Psi)_{AC}$, and by local discriminability we conclude that $(\mathcal{A} \otimes \mathcal{I}_C) | \Psi)_{AC} = (\mathcal{A}' \otimes \mathcal{I}_C) | \Psi)_{AC}$. Since this holds true for every $\Psi \in D_{\text{Ref}_\rho}$, we conclude that $\mathcal{A} =_{D_\rho} \mathcal{A}'$. The converse implication $\mathcal{A} =_{D_\rho} \mathcal{A}' \Rightarrow \mathcal{A} =_\rho \mathcal{A}'$ is trivial. ■

Based on the above identities of transformations we can also provide the notion of faithful state upon input of ρ and faithful state upon input of D_ρ (which again coincide for OPTs with local discriminability).

Definition 15 (Faithful state upon input of ρ). *A state $\Psi \in \text{St}(AB)$ is faithful upon input of $\rho \in \text{St}(A)$ if, given two arbitrary transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, C)$, the condition $\mathcal{A} | \Psi)_{AB} = \mathcal{A}' | \Psi)_{AB}$ implies that $\mathcal{A} =_\rho \mathcal{A}'$.*

Definition 16 (Faithful state upon input of D_ρ). *Given a system A and a state $\rho \in \text{St}(A)$, we say that a state $\Psi \in \text{St}(AB)$ is faithful upon input of D_ρ if, given two arbitrary transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, C)$, the condition $\mathcal{A} | \Psi)_{AB} = \mathcal{A}' | \Psi)_{AB}$ implies that $\mathcal{A} =_{D_\rho} \mathcal{A}'$.*

We simply say that a state $\Psi \in \text{St}(AB)$ is faithful for system A if, given two arbitrary transformations $\mathcal{A}, \mathcal{A}' \in \text{Transf}(A, C)$, the condition $\mathcal{A} | \Psi)_{AB} = \mathcal{A}' | \Psi)_{AB}$ implies that $\mathcal{A} = \mathcal{A}'$. Clearly a state $\Psi \in \text{St}(AB)$ that is faithful upon input of D_χ , with χ an internal state, is faithful for system A .

IV. INFORMATION AND DISTURBANCE

Within the general scenario of operational probabilistic theories, and without further assumptions on the structure of the theory, we aim at defining the notions of *non-disturbing* and *no-information tests*. These notions have already been investigated for theories that satisfy local discriminability or purification. We start highlighting the weakness of previous approaches in cases where the above hypotheses do not hold. The disturbance and the information produced by a test on a physical system A are commonly defined in relation to measurements and states of the system A only, disregarding the action of the same test on an enlarged systems AB .

A test $\{\mathcal{A}_i\}_{i \in X}$ on system A is said to be non-disturbing upon input of $\rho \in \text{St}(A)$ if for every σ in the refinement set of ρ it is $\sum_i \mathcal{A}_i |\sigma\rangle_A = |\sigma\rangle_A$, namely if $\sum_i \mathcal{A}_i =_\rho \mathcal{I}_A$ according to definition 12. However, this definition is not operationally consistent if applied to theories without local discriminability. A physically relevant example is that of the Fermionic theory [16] that, due to the parity super-selection rule, is non-local tomographic [17, 18] (it is 2-local tomographic according to Definition 9). We can see via a simple example that, for a Fermionic system A , a test $\{\mathcal{A}_i\}_{i \in X}$ such that $\sum_i \mathcal{A}_i =_\rho \mathcal{I}_A$ still can disturb the states of a composite system AB .

The parity superselection rule on a system N_F of N Fermions forbids any state corresponding to a superposition of vectors belonging to \mathcal{F}_N^e and \mathcal{F}_N^o , representing Fock vector spaces with total even and odd occupation numbers, respectively. As a consequence the set of states $\text{St}(N_F)$ splits in the direct sum of two spaces, containing the states with even and odd parity, respectively. It is now convenient to make use of the projectors onto the well-defined parity subspaces \mathbb{P}_e , for the even space, and \mathbb{P}_o , for the odd one. Notice that, since $\mathbb{P}_e \mathbb{P}_o = \mathbb{P}_o \mathbb{P}_e = 0$ any Fermionic state ρ will be of the form $\rho = \mathbb{P}_e \rho \mathbb{P}_e + \mathbb{P}_o \rho \mathbb{P}_o$. Consequently the *parity test* $\{\mathbb{P}_e \cdot \mathbb{P}_e, \mathbb{P}_o \cdot \mathbb{P}_o\}$ leaves every state ρ unchanged. Intuitively, this seems to suggest that parity can be measured without disturbing. Indeed, this view is in agreement with the notion of disturbance that has been considered in the literature so far.

Consider now a mixed state $\rho \in \text{St}(N_F)$, with $\rho = p_e \rho_e + p_o \rho_o$, ρ_e and ρ_o an even and an odd pure state respectively, and $p_e + p_o = 1$. For example, consider the states

$$\rho_e = |00\rangle\langle 00|, \quad \rho_o = |01\rangle\langle 01|,$$

and $p_e = p_o = 1/2$, so that

$$\rho = \frac{1}{2}(|00\rangle\langle 00| + |01\rangle\langle 01|).$$

Since Fermionic theory allows for purification [17], we can always find a state $\Psi \in \text{St}(M_F)$, with $M > N$ that purifies ρ . Since Ψ is pure, it has a definite parity, say even. In our example one can choose

$$\Psi = \frac{1}{2}(|000\rangle + |011\rangle)(\langle 000| + \langle 011|). \quad (4)$$

Therefore, the local test on the system N_F that measures the parity of the system, will not disturb the states of N_F but will decohere the state Ψ to a mixed state, then introducing a disturbance. For example, in our case

$$\begin{aligned} & (\mathbb{P}_e \otimes I)\Psi(\mathbb{P}_e \otimes I) + (\mathbb{P}_o \otimes I)\Psi(\mathbb{P}_o \otimes I) \\ &= \frac{1}{2}(|000\rangle\langle 000| + |011\rangle\langle 011|). \end{aligned}$$

In order to avoid the above issue, and to introduce a definition of non-disturbing test that works also for theories without local discriminability, one could say that a test $\{\mathcal{A}_i\}_{i \in X}$ on system A is non-disturbing upon input of $\rho \in \text{St}(A)$, if for every σ in the refinement set of ρ and every purification $\Psi_{AB} \in \text{St}(AB)$ of σ it is $\sum_i \mathcal{A}_i |\Psi\rangle_{AB} = |\Psi\rangle_{AB}$. This route, which has been proposed in Refs. [9, 10], captures the operational meaning of disturbance also for Fermionic systems. However, the definition of Refs. [9, 10] requires purification, and thus cannot be used in theories without purification, e. g. the cases of PR boxes, or the classical theory of information.

Based on the above motivations our proposal is to define the disturbance (and the information) produced by a test in terms of its action on dilations. This leads to notions of information and disturbance that are completely general and thus do not depend on local discriminability or purification. This will allow us to prove a no-information-without-disturbance theorem for a very large class of OPTs.

Definition 17 (Non-disturbing test upon input of D_ρ). *We say that a test $\{\mathcal{A}_i\}_{i \in X}$ on system A is non-disturbing upon input of D_ρ , with $\rho \in \text{St}(A)$, if*

$$\sum_i \mathcal{A}_i =_{D_\rho} \mathcal{I}_A. \quad (5)$$

Notice that, according to the above definition, the test $\{\mathcal{A}_i\}_{i \in X}$ is non-disturbing upon input of D_ρ if for every σ in the refinement set of ρ and for every dilation $\Psi_{AB} \in \text{St}(AB)$ of σ , it is

$$\sum_{i \in X} \mathcal{A}_i |\Psi\rangle_{AB} = |\Psi\rangle_{AB}. \quad (6)$$

This definition of disturbance thus stresses the effect of a transformation on correlations with remote systems.

We say that a test $\{\mathcal{A}_i\}_{i \in X}$ is *non-disturbing* if $\sum_i \mathcal{A}_i = \mathcal{I}_A$, namely the test is operationally identical

to the identity transformation of system A (it preserves any state Ψ_{AB} for every ancillary system B). In particular, this is the case if ρ in definition 17 is internal.

Clearly, a test on system A is disturbing upon input of D_ρ , with $\rho \in \text{St}(A)$ whenever it is not non-disturbing, namely there exists a σ in the refinement set of ρ and a dilation $\Psi_{AB} \in \text{St}(AB)$ of σ such that $\sum_i \mathcal{A}_i |\Psi\rangle_{AB} \neq |\Psi\rangle_{AB}$.

Remark 2. We could have defined a non-disturbing test from A to C upon input of D_ρ as follows

$$\sum_i \mathcal{A}_i =_{D_\rho} \mathcal{R}, \quad (7)$$

where $\mathcal{R} \in \text{Transf}(A, C)$ is a left-reversible transformation, namely there exists another transformation $\mathcal{W} \in \text{Transf}(C, A)$ such that $\mathcal{W}\mathcal{R} = \mathcal{I}_A$. However, the classification of non-disturbing tests according to this definition is trivially provided by the classification according to Definition 17. Indeed, the most general non-disturbing test from $A \rightarrow C$ is the sequence of tests of the form $\{\mathcal{A}_i \mathcal{R}\}_{i \in X}$, with $\{\mathcal{A}_i\}_{i \in X}$ non-disturbing according to Definition 17, and $\mathcal{R} \in \text{Transf}(A, C)$ left-reversible.

In the same spirit we can establish if a test provides information according to the following definition:

Definition 18 (No-information test upon input of D_ρ). *We say that a test $\{\mathcal{A}_i\}_{i \in X}$ with events $\mathcal{A}_i \in \text{Transf}(A, C)$ does not provide information on system A upon input of D_ρ , with $\rho \in \text{St}(A)$, if for every choice of deterministic effect $e \in \text{Eff}(CB)$ there exists a deterministic effect $f \in \text{Eff}(AB)$ such that*

$$(e|_{CB} \mathcal{A}_i =_{D_\rho} p_i(e) (f|_{AB}), \quad \forall i \in X. \quad (8)$$

Again we say that a test $\{\mathcal{A}_i\}_{i \in X}$ is a *no-information* test if for any choice of deterministic effects e_{CB} there exists a deterministic effect f_{AB} , such that

$$(e|_{CB} \mathcal{A}_i = p_i(e) (f|_{AB}), \quad \forall i \in X, \quad (9)$$

namely the occurrence probability of each outcome $i \in X$ does not depend on the state of the system. This is the case, e.g., if the test $\{\mathcal{A}_i\}_{i \in X}$ does not provide information upon input of D_ρ , with ρ internal state of the system.

Naturally a test $\{\mathcal{A}_i\}_{i \in X}$ provides information upon input of D_ρ , with $\rho \in \text{St}(A)$, whenever there exist states σ and σ' in the refinement set of ρ and $\Psi \in D_\sigma$, $\Psi' \in D_{\sigma'}$, with $\Psi, \Psi' \in \text{St}(AB)$, such that

$$\frac{(e|_{CB} \mathcal{A}_i | \Psi)_{AB}}{\sum_{j \in X} (e|_{CB} \mathcal{A}_j | \Psi)_{AB}} \neq \frac{(e|_{CB} \mathcal{A}_i | \Psi')_{AB}}{\sum_{j \in X} (e|_{CB} \mathcal{A}_j | \Psi')_{AB}}, \quad (10)$$

for some deterministic effect e_{CB} .

Remark 3. Notice that in Eq. (9) the probability of the transformation $\mathcal{A}_i \forall i \in X$ generally depends on the deterministic effect $(e|_{CB}$, this accounting for non causal

theories. In the more general case in which also the deterministic effect $(f|_{\text{AB}}$ on the right hand side of Eq. (9) depends on $i \in X$, the test $\{\mathcal{A}_i\}_{i \in X}$ would provide information on the system state (this will happen only for probabilistic states).

A. No-information without disturbance

In this section we state the condition of no-information without disturbance and introduce criteria for it to be satisfied by an OPT.

Definition 19 (No-information without disturbance upon input of D_ρ). *Consider an OPT and a state $\rho \in \text{St}(A)$, with A a system of the theory. Then the OPT satisfies no-information without disturbance upon input of D_ρ if for every test $\{\mathcal{A}_i\}_{i \in X} \subseteq \text{Transf}(A)$ that is non-disturbing upon input of D_ρ , the test does not provide information upon input of D_ρ .*

Definition 20 (OPT with no-information without disturbance). *We say that an OPT satisfies no-information without disturbance if it satisfies no-information without disturbance upon input of D_ρ for every $\rho \in \text{St}(A)$, for every system A .*

We prove now a necessary and sufficient condition for a theory to satisfy no-information without disturbance.

Theorem 1. *An OPT satisfies no-information without disturbance if and only if the identity transformation is atomic for every system of the theory.*

Proof. We start proving that if an OPT satisfies no-information without disturbance then the identity transformation is atomic. Consider a system A of the theory, and a refinement $\{\mathcal{A}_i\}_{i \in X}$ ($\mathcal{A}_i \in \text{Transf}(A)$ for every $i \in X$) of the identity map $\mathcal{I}_A = \sum_i \mathcal{A}_i$ for system A . The test $\{\mathcal{A}_i\}_{i \in X}$ is clearly non-disturbing, therefore by hypothesis it is a no-information test, namely for every deterministic effect $e \in \text{Eff}(AB)$, there exists a deterministic effect $f \in \text{Eff}(AB)$ such that for every $i \in X$ one has $(e|_{\text{AB}} \mathcal{A}_i = p_i(e)(f|_{\text{AB}}$. Summing both sides of the last equation over the index $i \in X$, and reminding that $\sum_{i \in X} p_i(e) = 1$, we find that $e = f$. Therefore, the no-information condition is

$$(e|_{\text{AB}} \mathcal{A}_i = p_i(e)(e|_{\text{AB}}, \quad (11)$$

for every deterministic effect $e \in \text{Eff}(AB)$. Consider now an arbitrary pair of pure states $\Psi_k \in \text{St}(AB)$, $k = 1, 2$. Since $\sum_i \mathcal{A}_i |\Psi_k\rangle_{\text{AB}} = |\Psi_k\rangle_{\text{AB}}$, it is

$$\mathcal{A}_i |\Psi_k\rangle_{\text{AB}} = \lambda_i(\Psi_k) |\Psi_k\rangle_{\text{AB}}, \quad \sum_i \lambda_i(\Psi_k) = 1, \quad (12)$$

where the coefficients $\lambda_i(\Psi_k)$ generally depend on the state Ψ_k . However, for each pure state Ψ_k there exists a deterministic effect $e_k \in \text{Eff}(AB)$ such that $(e_k | \Psi_k) \neq 0$.

Upon applying the deterministic effect $e := \frac{1}{2}(e_1 + e_2)$ on both sides of Eq. (12), we get

$$(e | \mathcal{A}_i |\Psi_k\rangle_{\text{AB}} = \lambda_i(\Psi_k) (e | \Psi_k)_{\text{AB}}. \quad (13)$$

Now, applying both sides of Eq. (11) to Ψ_k , we get

$$(e | \mathcal{A}_i |\Psi_k\rangle_{\text{AB}} = p_i(e) (e | \Psi_k)_{\text{AB}}, \quad (14)$$

and comparing the last two identities, considering that $(e | \Psi_k)_{\text{AB}} \neq 0$, we obtain

$$\lambda_i(\Psi_1) = \lambda_i(\Psi_2) = p_i(e), \quad \forall i \in X. \quad (15)$$

Since this holds true for every pair of pure states Ψ_k , we conclude that $\lambda_i(\Psi)$ is independent of Ψ . Moreover, by the same argument $p_i(e) \equiv p_i$ is independent of e . Then $\mathcal{A}_i |\rho\rangle_{\text{AB}} = p_i |\rho\rangle_{\text{AB}}$, $\forall \rho \in \text{St}(AB)$, proving that $\mathcal{A}_i = p_i \mathcal{I}_A$. Notice that we implicitly assumed that the probabilities p_i do not depend on the choice of the system B . Actually this can be proven as shown in Appendix B.

The converse implication, namely that if in a OPT the identity transformation is atomic then a non-disturbing test cannot provide information, is trivial. ■

The above theorem straightforwardly generalises upon input of D_ρ as follows.

Theorem 2. *Consider an OPT and a state $\rho \in \text{St}(A)$. Then the OPT satisfies no-information without disturbance upon input of D_ρ if and only if the identity \mathcal{I}_A is atomic upon input of D_ρ .*

The proof can be obtained from that of Theorem 1 by referring to Definitions 17 and 18, and substituting the pertaining equalities with those upon input of D_ρ .

Proposition 4. *An OPT satisfies no-information without disturbance if and only if for every system there exists an atomic transformation which is either left- or right-reversible.*

Proof. We start proving that a theory with an atomic reversible transformation for each system satisfies no-information without disturbance. Let $\mathcal{R} \in \text{Transf}(A, C)$ be atomic and left-reversible (the right-reversible case is analogous). Then consider a refinement $\mathcal{I}_A = \sum_i \mathcal{A}_i$, with $\mathcal{A}_i \in \text{Transf}(A)$ for $i \in X$, of the identity transformation. By definition of identity map it is $\mathcal{R} \mathcal{I}_A = \sum_i \mathcal{R} \mathcal{A}_i = \mathcal{R}$, and due to the atomicity of \mathcal{R} it must be $\mathcal{R} \mathcal{A}_i \propto \mathcal{R}$ for every $i \in X$. Since \mathcal{R} is left-reversible (namely there exists $\mathcal{W} \in \text{Transf}(C, A)$ such that $\mathcal{W} \mathcal{R} = \mathcal{I}_A$) it follows that $\mathcal{A}_i \propto \mathcal{I}_A$ for every $i \in X$, which proves the atomicity of \mathcal{I}_A .

The other implication, that in a theory that satisfies no-information without disturbance for every system there exists an atomic transformation which is either left- or right-reversible, is trivial. Indeed, in a theory that satisfies no-information without disturbance the identity, which is both right- and left-reversible, is atomic as proved in Theorem 1. ■

Proposition 5. *An OPT satisfies no-information without disturbance if and only if for every system every reversible transformation is atomic.*

Proof. We prove that if the theory satisfies no-information without disturbance, then every reversible transformation is atomic. Indeed, let $\mathcal{R} \in \text{Transf}(A)$ be reversible, and suppose that $\mathcal{R} = \sum_{i \in X} \mathcal{R}_i$ for test $\{\mathcal{R}_i\}_{i \in X}$. Then, one has

$$\sum_{i \in X} \mathcal{R}_i \mathcal{R}^{-1} = \mathcal{I}_A, \quad (16)$$

and by theorem 1 one has that $\mathcal{R}_i \mathcal{R}^{-1} = p_i \mathcal{I}_A$. Finally, multiplying by \mathcal{R} to the right, we conclude that $\mathcal{R}_i = p_i \mathcal{R}$, namely the refinements of \mathcal{R} must be trivial. For the converse, it is sufficient to observe that the identity is reversible. ■

B. Information without disturbance

In this section we provide the general structure of the state space of any theory where some information can be extracted from a system without introducing disturbance. Such information is “classical” in the sense that the measurement is the reading of information that is repeatable and sharable. In particular, for the classical OPT the whole information encoded on a system can be read in this way. The proof of the above statements are based on the following theorem.

Theorem 3. *The non redundant atomic refinement of the identity is unique for every system. Moreover, given the non redundant atomic refinement $\{\mathcal{A}_i\}_{i \in X} \subseteq \text{Transf}(A)$ of the identity $\mathcal{I}_A = \sum_i \mathcal{A}_i$, one has $\mathcal{A}_i \mathcal{A}_j = \mathcal{A}_i \delta_{ij}$.*

Proof. Suppose that the identity transformation of system A allows for two atomic refinements $\mathcal{I}_A = \sum_{i \in X} \mathcal{A}_i$, and $\mathcal{I}_A = \sum_{j \in Y} \mathcal{B}_j$. Since $\sum_i \mathcal{A}_i \mathcal{B}_j = \mathcal{B}_j$, from the atomicity of the transformations \mathcal{B}_j we get $\mathcal{A}_i \mathcal{B}_j = c_{ij} \mathcal{B}_j$, for some $c_{ij} \geq 0$ such that $\sum_{i \in X} c_{ij} = 1 \forall j \in Y$. Similarly we get $\mathcal{A}_i \mathcal{B}_j = d_{ij} \mathcal{A}_i$ for some $d_{ij} \geq 0$ such that $\sum_{j \in Y} d_{ij} = 1 \forall i \in X$. Then $c_{ij} \mathcal{B}_j = d_{ij} \mathcal{A}_i$. By non redundancy one has that for fixed j there is only one value of $i = i(j)$ such that $c_{ij} > 0$, and normalisation gives $c_{i(j)j} = 1$. By a similar argument for a fixed i there is $j = j(i)$ such that $d_{ij(i)} = 1$. Then one has $\mathcal{B}_j = \mathcal{A}_{i(j)}$. This proves uniqueness of the non redundant atomic refinement of the identity.

By the same argument as before, for the non redundant atomic refinement of the identity one has $\mathcal{A}_i \mathcal{A}_j = c_{ij} \mathcal{A}_j = d_{ij} \mathcal{A}_i$, for some $c_{ij}, d_{ij} \geq 0$ such that $\sum_{i \in X} c_{ij} = \sum_{j \in X} d_{ij} = 1 \forall i, j \in X$. By atomicity and non redundancy one must have $c_{ij} = d_{ij} = \delta_{ij}$. ■

The above theorem has as a consequence the following structure theorem for OPTs.

Corollary 1. *For any pair of systems A, B of an OPT one has the following decomposition of the set of states of AB*

$$\text{St}(AB) = \bigoplus_{(i,j) \in X \times Y} \text{St}_{ij}(AB), \quad (17)$$

where for non redundant atomic decompositions $\{\mathcal{A}_i\}_{i \in X}$, $\{\mathcal{B}_j\}_{j \in Y}$ of the identities \mathcal{I}_A and \mathcal{I}_B , one has

$$(\mathcal{A}_i \otimes \mathcal{B}_j) | \Psi_{i'j'} \rangle = \delta_{ii'} \delta_{jj'} | \Psi_{ij} \rangle, \quad (18)$$

for all $\Psi_{i'j'} \in \text{St}_{i'j'}(AB)$.

Remark 4. Notice that from Eq. (17) it trivially follows that for any system the block decomposition holds

$$\text{St}(A) = \bigoplus_{i \in X} \text{St}_i(A). \quad (19)$$

However, Eq. (17) contains the information that the decomposition holds in that specific form also for composite systems. This is not a straightforward consequence of the decomposition of local states, as witnessed by the fermionic case. Indeed, the state in Eq. (4) does not have definite parity for the two subsystems corresponding to two fermions on the left and one on the right, hence the state space cannot be of the form in Eq. (17).

Remark 5. For a theory without atomicity of parallel composition one has also the possibility that the refinement $\mathcal{A}_i \otimes \mathcal{B}_j$ in Eq. (18) of \mathcal{I}_{AB} is not atomic. In such case one has $\text{St}(AB) = \bigoplus_{k \in Z} \text{St}_k(AB)$, and $\text{St}_{ij}(AB) = \bigoplus_{k \in Z_{ij}} \text{St}_k(AB)$, for some partition Z_{ij} of Z .

In the following Corollary we formalise the fact that a theory where any information can be extracted without disturbance must have classical sets of states.

Corollary 2. *If an OPT is such that it does not satisfy no-information without disturbance upon input of D_ρ , for every state $\rho \in \text{St}(A)$ of every system A, then every system of the theory is classical.*

Proof. The proof follows from the simple fact that all blocks $\text{St}_{ij}(AB)$ in Corollary 1 are one-dimensional, i.e. all pure states are jointly perfectly discriminable, namely all the systems of the theory are classical. ■

Remark 6. We remind that a system is classical when all its pure states are jointly perfectly discriminable. As a consequence the set of states of a classical system is a simplex. A special case of theory whose systems are all classical is the usual classical information theory. However, even when all systems are classical, the theory can be non classical because the system composition does not satisfy local discriminability (see Ref. [26]).

C. Sufficient conditions for no-information without disturbance

In this section we prove some further conditions for no-information without disturbance that are only sufficient. The first condition is expressed in the following proposition and in its corollary.

Proposition 6. *An OPT satisfies no-information without disturbance upon input of D_ρ , with $\rho \in \text{St}(A)$, if there exists a pure state $\Psi \in D_{\text{Ref}_\rho}$ that is faithful upon input of D_ρ .*

Proof. Given a system A and a state $\rho \in \text{St}(A)$, let $\Psi \in D_{\text{Ref}_\rho}$ be pure and faithful upon input of D_ρ (see Definition 16). Now let the test $\{\mathcal{A}_i\}_{i \in X} \in \text{Transf}(A)$ be non-disturbing upon input of D_ρ , namely $\sum_i \mathcal{A}_i =_{D_\rho} \mathcal{I}_A$. Then, since $\Psi \in D_{\text{Ref}_\rho}$ we have $\sum_i \mathcal{A}_i |\Psi\rangle = |\Psi\rangle$, and since Ψ is pure, there exists a set of probabilities $\{p_i\}_{i \in X}$ such that $\mathcal{A}_i |\Psi\rangle = p_i |\Psi\rangle$. However, due to the faithfulness of Ψ , the map $\mathcal{A} \mapsto \mathcal{A} |\Psi\rangle$ is injective upon input of D_ρ , and we conclude that $\mathcal{A}_i =_{D_\rho} p_i \mathcal{I}_A$, and, by definition, the test $\{\mathcal{A}_i\}_{i \in X} \in \text{Transf}(A)$ does not extract information upon input of D_ρ . ■

Corollary 3. *An OPT satisfies no-information without disturbance if for every system A there exists a pure faithful state.*

In the next proposition we show that the OPTs that satisfy purification are a subset of the OPTs with no-information without disturbance. In the following we will see that it is actually a proper subset (see also Fig. 1).

Proposition 7. *An OPT with purification, satisfies no-information without disturbance.*

Proof. Given an OPT with purification suppose that it violates the no-information without disturbance, namely there exists a system A such that \mathcal{I}_A is not atomic. Then let $\mathcal{I}_A = \sum_i \mathcal{A}_i$, for some atomic non redundant test $\{\mathcal{A}_i\}_{i \in X} \subseteq \text{Transf}(A)$. Let us consider a mixed state $\text{St}(A) \ni |\rho\rangle = \sum_{i \in X} p_i |\sigma_i\rangle$ with $p_i |\sigma_i\rangle := \mathcal{A}_i |\rho\rangle$, and $\{p_i\}_{i \in X}$ a probability distribution with $p_i > 0 \forall i$. Then by Theorem 3 we have $\mathcal{A}_i |\sigma_j\rangle = \delta_{ij} |\sigma_i\rangle$. Since the theory allows for purification, let $\Psi \in \text{St}(AB)$ be a purification of ρ for deterministic effect $e \in \text{Eff}(B)$. Now, on one hand since the test $\{\mathcal{A}_i\}_{i \in X}$ refines the identity, it is $|\Psi\rangle_{AB} = \sum_{i \in X} \mathcal{A}_i |\Psi\rangle_{AB}$, and being Ψ pure it must be $\mathcal{A}_i |\Psi\rangle_{AB} = q_i |\Psi\rangle_{AB}$, with $\{q_i\}_{i \in X}$ a probability distribution. On the other hand, for every $i \neq j$ the marginals with deterministic effect $e \in \text{Eff}(B)$ of $\mathcal{A}_i |\Psi\rangle_{AB}$ and $\mathcal{A}_j |\Psi\rangle_{AB}$ are perfectly discriminable. But this contradicts the fact that $\mathcal{A}_i |\Psi\rangle_{AB}$ and $\mathcal{A}_j |\Psi\rangle_{AB}$ are both proportional to Ψ . ■

V. OUTLOOK ON NO-INFORMATION WITHOUT DISTURBANCE

In this last section we analyse the relation between no-information without disturbance and other proper-

ties of operational probabilistic theories. Here we focus on local discriminability and purification that, being typical quantum features, are commonly associated with no-information without disturbance. Here, via concrete examples, we show that no-information without disturbance can actually be satisfied independently of the above two properties.

The following proposition proves that the probabilistic theory [27–31] corresponding to the PR-boxes model, originally introduced in Ref. [32], satisfies no-information without disturbance.

Proposition 8. *The PR-boxes theory satisfies no-information without disturbance.*

Proof. This can be proved in several ways. For example we show that any system of the theory allows for a reversible atomic transformation and then use Proposition 4. The fact that any system has a reversible atomic transformation follows from the following three points. I) The reversible transformations of the elementary system A of the theory (the convex set of normalized states of A is represented by a square, and the set of reversible transformations of A coincides with the set of symmetries of a square, the dihedral group of order eight D_8 , containing four rotations and four reflections) are atomic [29]. II) From Refs. [33, 34] we know that the set of reversible maps of the N -partite system $A^{\otimes N}$ is generated by local reversible operations plus permutations of the systems. Accordingly, the system $A^{\otimes N}$ allows for a multipartite reversible transformation $U_1 \otimes U_2 \otimes U_N$ made of local reversible transformations U_i , $i = 1, \dots, N$. III) Since PR-boxes satisfy local discriminability, the chosen multipartite transformation is atomic due to the atomicity of parallel composition (see Proposition 2). ■

The last proposition leads to the following relevant corollary.

Corollary 4. *No-information without disturbance does not imply purification.*

Proof. PR-boxes theory with minimal tensor product [14], which satisfies no-information without disturbance (see the above Proposition 8), indeed does not satisfy purification, since every pure state is a tensor product of local pure states. ■

Turning to the case of local discriminability, we now show that it is independent on no-information without disturbance. Indeed, there are theories that satisfy the former and not the latter (e.g. classical theory), and viceversa, theories that satisfy the latter but not the former, e.g. the fermionic [16–18] and the real [17, 18, 23] quantum theories. This follows from Proposition 7, considering that both theories satisfy purification. Therefore we have the following corollary.

Corollary 5. *No-information without disturbance and local discriminability are independent.*

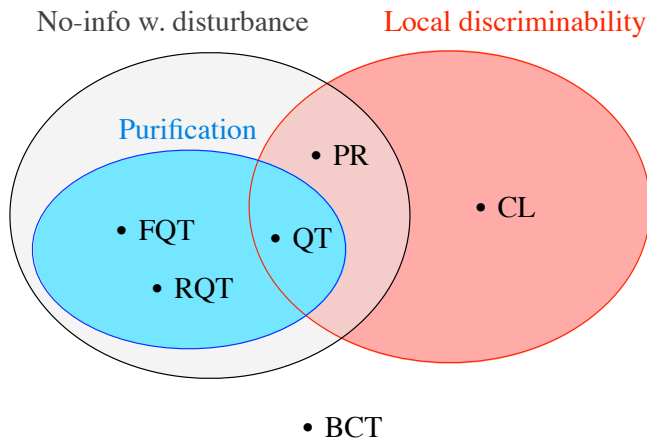


Figure 1. Comparing OPTs that satisfy no-information without disturbance (grey set), local discriminability (red set) and purification (blue set). Quantum theory (QT) lies at the intersection of the three sets. The set of OPTs with purification is a proper subset of OPTs with no-information without disturbance. An example of OPT that satisfies no-information without disturbance but violates purification is the PR-boxes theory with minimal tensor product (PR). Moreover, PR-boxes satisfy local discriminability, providing a non-trivial intersection between local discriminability and no-information without disturbance in the absence of purification. We observe that no-information without disturbance is independent of local discriminability and viceversa. Indeed classical theory (CL) satisfies only local discriminability while Fermionic quantum theory (FQT) and real quantum theory (RQT) satisfy only no-information without disturbance. Finally, it has been shown in Ref. [24] that there exist OPTs without local discriminability, that have all systems classical, thus retaining the possibility of extracting all the information without disturbance. An example is the *bilocal classical theory* (BCT) of the same Ref. [24], which satisfies 2-local discriminability (see Definiton 9).

Finally, we observe that as a consequence of Corollary 2, and subsequent remark, the classical theory of information is the only theory with local discriminability in which all the information can be extracted without disturbance. However, in the absence of local discriminability, it is still possible to have other theories where all the information can be extracted without disturbance. This has been proved in Ref. [24] where the authors describe an OPT whose systems of any dimension are classical (and then violate no-information without disturbance), but with a parallel composition that differs from the usual classical one, leading to a violation of local discriminability, more precisely to a 2-local theory according to Definition 9.

VI. CONCLUSIONS

We have analysed the interplay between information and disturbance for a general operational probabilistic theory, considering the effect of measurements also on entanglement with the environment, differently from the traditional approach focused only on the measured system. Indeed, the two resulting notions of disturbance coincide only in special cases, such as quantum theory, as well as every theory that satisfies local discriminability. Our approach is universal for any OPT, including theories without causality or purification. In this setting we proved that the atomicity of the identity transformation is an equivalent condition for no-information without disturbance.

We have characterized the structure of theories where the identity is not atomic, showing that in this case the information that can be extracted without disturbance is “classical”, in the sense that it is sharable and repeatable. On the other hand, the typical situation for a general OPT entails information whose extraction requires disturbance—the only exception being classical theory.

While no information without disturbance is a consequence of purification, the converse is not true, as pointed out by counter-examples. Similarly we have shown that no-information without disturbance and local discriminability are independent properties of the theory.

Our results are expected to have immediate applicability to secure key-distribution. Indeed, a physical theory including a system (or even just a set of states of a system) that satisfies no-information without disturbance can guarantee a private and reliable channel for distributing messages. The idea of studying secure key-distribution in a framework more general than the classical and the quantum ones has been proposed in Refs. [12, 14], and the present generalisation of no-information without disturbance to arbitrary OPTs is a first step in proving that secure key-distribution is possible in any non-classical theory.

ACKNOWLEDGMENTS

This publication was made possible through the support of a grant from the John Templeton Foundation, ID No. 60609 “Quantum Causal Structures.” The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

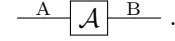
[1] W. Heisenberg, *Zeitschrift für Physik* **43**, 172 (1927).
 [2] P. Busch, T. Heinonen, and P. Lahti, *Physics Reports* **452**, 155 (2007).

[3] P. Busch, P. Lahti, and R. F. Werner, *Rev. Mod. Phys.* **86**, 1261 (2014).
 [4] C. A. Fuchs and A. Peres, *Phys. Rev. A* **53**, 2038 (1996).

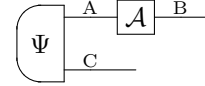
- [5] G. M. D’Ariano, Fortschritte der Physik: Progress of Physics **51**, 318 (2003).
- [6] M. Ozawa, Annals of Physics **311**, 350 (2004).
- [7] L. Maccone, Phys. Rev. A **73**, 042307 (2006).
- [8] P. Busch, ““no information without disturbance”: Quantum limitations of measurement,” in *Quantum Reality, Relativistic Causality, and Closing the Epistemic Circle: Essays in Honour of Abner Shimony* (Springer Netherlands, Dordrecht, 2009) pp. 229–256.
- [9] G. M. D’Ariano, G. Chiribella, and P. Perinotti, *Quantum Theory from First Principles: An Informational Approach* (Cambridge University Press, 2017).
- [10] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Phys. Rev. A **81**, 062348 (2010).
- [11] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Quantum from principles,” in *Quantum Theory: Informational Foundations and Foils*, edited by G. Chiribella and R. W. Spekkens (Springer Netherlands, Dordrecht, 2016) pp. 171–221.
- [12] J. Barrett, Physical Review A **75**, 032304 (2007).
- [13] G. Kimura, K. Nuida, and H. Imai, Reports on Mathematical Physics **66**, 175 (2010).
- [14] H. Barnum and A. Wilce, Electronic Notes in Theoretical Computer Science **270**, 3 (2011), proceedings of the Joint 5th International Workshop on Quantum Physics and Logic and 4th Workshop on Developments in Computational Models (QPL/DCM 2008).
- [15] T. Heinosaari, L. Leppäjärvi, and M. Plávala, Quantum **3**, 157 (2019).
- [16] S. B. Bravyi and A. Y. Kitaev, Annals of Physics **298**, 210 (2002).
- [17] G. M. D’Ariano, F. Manessi, P. Perinotti, and A. Tosini, International Journal of Modern Physics A **29**, 1430025 (2014).
- [18] G. M. D’Ariano, F. Manessi, P. Perinotti, and A. Tosini, EPL (Europhysics Letters) **107**, 20009 (2014).
- [19] We remind that for non causal theories the marginal is not unique, hence more generally, we require that *one of the marginals* is the given state.
- [20] C. Pfister and S. Wehner, Nature Communications **4**, 1851 EP (2013).
- [21] Notice that a more detailed account needs a category-theoretical definition of parallel and sequential composition of systems (see Ref. [11]).
- [22] G. M. D’Ariano, F. Manessi, and P. Perinotti, Physica Scripta **T163**, 014013 (2014).
- [23] L. Hardy and W. K. Wootters, Foundations of Physics **42**, 454 (2012).
- [24] G. Chiribella, G. D’Ariano, and P. Perinotti, Phys. Rev. A **84**, 012311 (2011).
- [25] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, 2010).
- [26] G. M. D’Ariano, M. Erba, P. Perinotti, in preparation.
- [27] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
- [28] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Phys. Rev. A **71**, 022101 (2005).
- [29] G. M. D’Ariano and A. Tosini, Quantum Information Processing **9**, 95 (2010).
- [30] A. J. Short and J. Barrett, New Journal of Physics **12**, 033034 (2010).
- [31] M. Dall’Arno, S. Brandsen, A. Tosini, F. Buscemi, and V. Vedral, Phys. Rev. Lett. **119**, 020401 (2017).
- [32] S. Popescu and D. Rohrlich, Foundations of Physics **24**, 379 (1994).
- [33] D. Gross, M. Müller, R. Colbeck, and O. C. O. Dahlsten, Phys. Rev. Lett. **104**, 080402 (2010).
- [34] S. W. Al-Safi and A. J. Short, Journal of Physics A: Mathematical and Theoretical **47**, 325303 (2014).

Appendix A

In the operational framework any event \mathcal{A} induces a map between states. Consider for example the event



For every choice of ancillary system C, and for every state $\Psi \in \text{St}(AC)$, the event \mathcal{A} maps the state Ψ to the state given by the following circuit



Accordingly, while states and effects are linear functionals over each other, we can always look at an event as a map between states

$$\mathcal{A} : |\Psi\rangle_{AC} \in \text{St}(AC) \mapsto \mathcal{A}|\Psi\rangle_{AC} \in \text{St}(BC), \quad (\text{A1})$$

(and similarly as a map between effects, from $\text{Eff}(AC)$ to $\text{Eff}(BC)$). The map above can be linearly extended to a map from $\text{St}_{\mathbb{R}}(AC)$ to $\text{St}_{\mathbb{R}}(BC)$ (we denote the extended map with the same symbol \mathcal{A}) and this extension is unique. Indeed a linear combination of states of AC is null, say $\sum_i c_i |\Psi_i\rangle$, if and only if $\sum_i c_i (a|\Psi_i) = 0$ for any $a \in \text{Eff}(BC)$. Moreover, since $(b|_B \mathcal{A} \in \text{Eff}(AC)$ for every $b \in \text{Eff}(B)$, then $\forall b \in \text{Eff}(B)$ it is $0 = (b|_B \mathcal{A}(\sum_i c_i |\Psi_i)_{AB}) = \sum_i c_i (b|_B \mathcal{A}|\Psi_i)_{AB} = (b|_B \sum_i c_i \mathcal{A}|\Psi_i)_{AB}$, and we finally get $\sum_i c_i \mathcal{A}|\Psi_i)_{AB} = 0$.

Appendix B

Given a transformation $\mathcal{A} \in \text{Transf}(A)$ such that for every system B there exists p_B such that

$$\mathcal{A}|\rho\rangle_{AB} = p_B |\rho\rangle_{AB}, \quad (\text{B1})$$

then actually $p_B \equiv p$ cannot depend on the system B. Indeed, choosing $\rho = \tau \otimes \sigma_B$ with arbitrary $\tau \in \text{St}(A)$ and normalised $\sigma \in \text{St}(B)$, and discarding system B on both sides of Eq. (B1) $\forall \rho \in \text{St}(AB)$, one obtains $\mathcal{A}|\tau\rangle = p_B |\tau\rangle$. This clearly shows that $p_B \equiv p$.