Equivalence between Markovianity and monotonic decrease of information

Francesco Buscemi\textsuperscript{1} and Nilanjana Datta\textsuperscript{2}\textsuperscript{1}

\textsuperscript{1}Graduate School of Information Science, Nagoya University, Chikusa-ku, Nagoya, 464-8601, Japan
\textsuperscript{2}Statistical Laboratory, University of Cambridge, Cambridge CB3 0WB, U.K.

The crucial feature of a Markov chain is its memoryless property. This, in particular, implies that any information about the state of a physical system, whose evolution is governed by a Markov chain, can only decrease as the system evolves. Here we show that such a decrease of information is not only necessary but also sufficient for a discrete-time stochastic evolution to be Markovian. Moreover, this condition is valid both in the quantum and classical settings.

Discrete-time Markov chains are stochastic processes whose state at each discrete instant in time \( t_i \) depends only on the state at the previous instant \( t_{i-1} \) and not on those at earlier times \( t_{i-2}, t_{i-3}, \ldots, t_0 \); see, e.g., Refs. \[20\] and \[19\]. This fact makes Markov chains the natural candidates to model memoryless processes. As many real-world situations seem to be memoryless (at least in a first approximation), Markov chains are ubiquitous in many fields, ranging from physics, chemistry, biology, information and computer sciences, to economics and social sciences.

A direct consequence of the lack of memory is that the information about the initial state of the Markov chain cannot increase under dynamic evolution: the information lost at some point during the evolution cannot be injected back into the chain at later times, as this would imply the existence of some sort of memory. We refer to this property of Markovian evolution as the data-processing principle. This is because it can be formulated in terms of various inequalities (often called data-processing inequalities) which are satisfied by certain information-theoretic quantities (such as the mutual information \[10\]) evaluated at successive instants in time, as the Markov chain evolves.

This principle is also valid in the quantum setting, even though it is known that the classical definition of a Markov chain cannot be directly extended to the quantum domain. The problem with such an extension arises from the following fact. In defining a discrete-time Markov chain \( \{X_n; n \in \mathbb{N}\} \) (where the \( X_n \)'s are discrete random variables and the subscript \( n \) labels the instants in time) one starts with the joint probability distribution \( \Pr(X_0 = x_0, X_1 = x_1, \ldots, X_N = x_N) = p(x_0, x_1, \ldots, x_N) \) describing the stochastic process \[10\]. However, a quantum stochastic process does not admit such a description, since quantum theory is inherently non-commutative \[22, 30\]. In spite of this, there is a simple way to arrive at a description of a quantum Markov chain starting from a classical one. To see this, note that a classical stochastic process is Markovian if and only if the joint probability distribution can be factorized as \( p(x_0)p(x_1|x_0) \cdots p(x_N|x_{N-1}) \). The conditional probabilities in this expression can be considered to arise from successive actions of a sequence of independent channels which transform the initial state \( x_0 \) of the Markov chain (which occurs with probability \( p(x_0) \)) to the successive states \( x_1, x_2, \ldots \). This is because a classical channel is represented by a set of conditional probabilities. Hence, for any given initial state, a discrete-time classical stochastic process is Markovian if the final state of the system at any later instant in time can be obtained from the initial state by the successive actions of a sequence of classical channels at the intermediate times.

A quantum analogue of such a process is one in which the initial state and classical channels are respectively replaced by a density matrix and quantum channels (i.e., linear, completely-positive trace-preserving maps).

The advantage of such an analogy is twofold: on one hand, it concretely captures the idea of the lack of memory by stipulating that the system's state at each discrete instant in time only depends on the state in the immediately preceding instant, via the action of a channel; on the other hand, it allows us to treat classical and quantum Markov chains on the same footing.

The fact that discrete-time Markov chains necessarily obey the data-processing principle suggests that an information-theoretic criterion for Markovianity can be stated in terms of data-processing inequalities. In fact, the latter serve as witnesses of non-Markovian since the violation of any such inequality implies that the underlying stochastic process is non-Markovian (see, in particular, Refs. \[2, 9, 12, 13, 17, 23, 25, 29–31\] and the recent comprehensive review by Rivas, Huelga and Plenio \[22\]). However, each of the information-theoretic criteria for Markovianity obtained thus far have the drawback of being necessary but not sufficient \[9, 22\]. In contrast, in this paper we provide an information-theoretic condition for Markovianity which is both necessary and sufficient. We thus establish the first, fully general information-theoretic underpinning of Markovianity, which is valid both in the classical and quantum setting. It states the equivalence between Markovianity and the impossibility of increase of information under dynamic evolution. We prove this equivalence by showing that, amongst the many different inequalities in terms of which the data-processing principle can be formulated, there exists one which, if never violated, guarantees that the underlying stochastic process is Markovian. In other words, we ex-
plicitly formulate a data-processing inequality which is necessarily violated whenever the underlying stochastic process is not Markovian.

Definitions.—For the sake of generality, we start with the case of discrete-time quantum stochastic processes, recovering their classical analogues under suitable extra assumptions. In what follows, we only consider quantum systems defined on finite dimensional Hilbert spaces $\mathcal{H}$. The definitions used here closely adhere to those given in standard textbooks [18, 21]. We denote by $L(\mathcal{H})$ the set of all linear operators acting on $\mathcal{H}$, and by $D(\mathcal{H})$ the set of all density operators (or states) $\rho \in L(\mathcal{H})$, with $\rho \geq 0$ and $\text{Tr}[\rho] = 1$. The identity operator in $L(\mathcal{H})$ is denoted by the symbol $I$. An ensemble $\mathcal{E} = \{p(x)\rho^x\}_{x \in \mathcal{X}}$ is a finite family of states $\rho^x$, and a priori probabilities $p(x)$, labeled by the elements of a finite set $\mathcal{X}$. A positive-operator valued measure (POVM) is a finite family of positive semi-definite operators $\{P_x\}_{x \in \mathcal{Y}}$ such that $\sum_{x \in \mathcal{Y}} P_x = I$, labeled be the elements of a finite set $\mathcal{Y}$. A quantum channel is a linear, completely positive trace-preserving (CPTP) map $N : L(H_A) \rightarrow L(H_B)$. The identity channel from $L(H)$ to itself is denoted by id.

The physical model we consider is that of a quantum system which, at an initial time $t_0$, is put in contact with its surrounding environment and allowed to evolve jointly with the latter through successive discrete instants in time $t_1, t_2, \ldots, t_N$. We suppose that the environment, at time $t_0$, is in some initial state $\sigma_E$, which is uncorrelated with the state of the system. Hence, if the initial state of the system is $\rho_S^0$, its state at time $t_i$ is given by

$$\rho_S^i = \text{Tr}_E[U_i (\rho_S^0 \otimes \sigma_E) U_i^\dagger],$$

(1)

where the $U_i$’s are the unitary operators modeling the joint system-environment evolution from $t_0$ to $t_i$. We stress that the unitary operators in Eq. (1) can be arbitrary. In particular, our analysis can be easily generalized to the case in which the Hilbert spaces associated with the system and the environment vary with time, i.e., $\mathcal{H}_S$ and $\mathcal{H}_E$: in such a case, the unitary operators $U_i$ become isometries $V_i : \mathcal{H}_S \otimes \mathcal{H}_E \rightarrow \mathcal{H}_S \otimes \mathcal{H}_E$, with $V_i^\dagger V_i = I$, for all $i$. For the sake of simplicity we avoid the consideration of such a time-varying scenario in this paper. It is, however, considered in the Supplemental Material.

In this paper we assume, in particular, that the system’s state in (1) can be arbitrarily initialized, so that Eq. (1) can be used to define a finite sequence of quantum channels $\mathcal{N} = \{N^i : L(\mathcal{H}_S) \rightarrow L(\mathcal{H}_S) : 1 \leq i \leq N\}$, whose $i$-th element is defined by

$$N^i(\rho_S^0) := \text{Tr}_E[U_i (\rho_S^0 \otimes \sigma_E) U_i^\dagger].$$

In the above formula, each quantum channel $N^i$ propagates any given state $\rho_S^0 \in D(\mathcal{H}_S)$ of the quantum system at an initial time $t_0$ to its state at time $t_i$; see Figure 1 for a schematic representation. In fact, the Stinespring-Kraus representation theorem [10, 26] guarantees that such a sequence of channels is all we need to give an information-theoretic description of a discrete-time quantum stochastic process. This leads us to the following natural definition.

**Definition 1.** A discrete-time quantum stochastic process is given by an arbitrary sequence of quantum channels $\mathcal{N} = \{N^i : 1 \leq i \leq N\}$, where each $N^i : L(\mathcal{H}_S) \rightarrow L(\mathcal{H}_S)$ models the evolution of the state of a quantum system at an initial time $t_0$, to its state at a later time $t_i$. Further, such a process is Markovian if there exist $N-1$ quantum channels $C^i : L(\mathcal{H}_S) \rightarrow L(\mathcal{H}_S)$, $1 \leq i \leq N-1$, such that

$$N^{i+1} = C^i \circ N^i,$$

(2)

for all $i$ between 1 and $N-1$. See Fig. 1 for a schematic representation.

**The data-processing principle.**—As we mentioned before, the fact that information can only decrease along a Markov chain can be formalized in many ways, via a number of data-processing inequalities [11]. In the following we focus on one such data-processing inequality which has the particular advantage of being simple and yet operationally relevant.

Suppose that the experimenter knows a priori that the system is initially in some state $\rho_S^0 \in D(\mathcal{H}_S)$ with probability $p(x)$. Hence, the experimenter’s initial knowledge about the system is modeled by an ensemble $\mathcal{E} = \{p(x)\rho_S^0\}_{x}$. The question is: how can one quantify the information possessed by the experimenter at this point? Clearly such information depends on the distinguishability of the states in the ensemble: higher the distinguishability of the states, more is the information available to the experimenter. A natural measure of the information
about the system’s initial state is therefore given by the
guessing probability \[ P_{\text{guess}}(\mathcal{E}) := \max_x \sum_p(x) \text{Tr}[P^x_S \rho^x_S], \]
where the maximization is over all POVMs \( \{P^x_S\}_x \) on \( \mathcal{H}_S \). The fact that the guessing probability cannot in-
crease under the action of a channel on the states of the
ensemble is a very simple consequence of its definition.

**Definition 1.** A given discrete-time quantum stochastic
process \( \mathcal{F} = \{N^i; 1 \leq i \leq N\} \) is said to be
information decreasing if and only if, for any ensemble \( \mathcal{E} = \{p(x); \rho^x_S\}_x \), the sequence of guessing probabilities
\( [P_{\text{guess}}(\mathcal{E}_i)]_{i \in \mathbb{N}} \), where \( \mathcal{E}_i := \{p(x); N^i(\rho^x_S)\}_x \), is mono-
tonically non-increasing, i.e., \( P_{\text{guess}}(\mathcal{E}_i) \geq P_{\text{guess}}(\mathcal{E}_{i+1}) \)
for all \( 1 \leq i \leq N - 1 \).

The above definition constitutes our formalization of the
fact that information about the initial state of the
system cannot increase as the system evolves. This, of
course, has to happen irrespective of the information
about the system that the experimenter initially has.
This fact that is reflected in the above definition by the
requirement that the guessing probability cannot increase
for any ensemble of initial states, i.e., for any finite set \( \mathcal{F} \),
for any probability distribution on \( \mathcal{F} \), and for any
collection of states \( \rho^x_S \in \mathcal{D}(\mathcal{H}_S) \).

This paper builds upon a series of results extending
the so-called Blackwell-Sherman-Stein theorem \[11, 27, 28\]
of classical statistics to quantum statistical decision theory.
In particular, a crucial role in this paper is played by the
following result, recently proved in Ref. [6].

**Proposition 1.** A given discrete-time quantum stochastic
process \( \mathcal{F} = \{N^i; 1 \leq i \leq N\} \) is said to be
information decreasing if and only if, for all finite en-
sembles \( \mathcal{E} = \{p(x); \rho^x_S\}_x \) of states on \( \mathcal{H}_S \),
the following are equivalent:

1. \( P_{\text{guess}}(\mathcal{E}_i) \geq P_{\text{guess}}(\mathcal{E}_{i+1}) \) for
   any finite ensemble \( \mathcal{E}_i = \{p(x); N^i(\rho^x_S)\}_x \),
   where \( \mathcal{E}_i := \{p(x); N^i(\rho^x_S)\}_x \).
2. For any POVM \( \{Q^y\}_y \), there exists a correspond-
ing POVM \( \{P^y\}_y \) such that
   \( \text{Tr}[N^i(\rho_S) \; P^y] = \text{Tr}[N^2(\rho_S) \; Q^y] \),
   for any \( y \in \mathcal{F} \) and any \( \rho_S \in \mathcal{D}(\mathcal{H}_S) \).

In other words, the guessing probability for a channel
\( N^1 \) (for any ensemble) is greater than or equal to that
for a channel \( N^2 \) if and only if the image of \( N^2 \) (in the
Heisenberg picture) is contained in that of \( N^1 \). We
use the notation \( N^i \succeq N^2 \) (which denotes a partial order-
ing between channels) whenever one of the above condi-
tions holds. Accordingly, a given discrete-time quantum
stochastic process \( \mathcal{F} = \{N^i; 1 \leq i \leq N\} \) is information
decreasing if and only if \( N^1 \succeq N^2 \succeq \cdots \succeq N^N \).

**The quantum case.**—We now turn to the case in which
the discrete-time stochastic process is fully quantum, i.e.,
the channels in \( \mathcal{F} = \{N^i; 1 \leq i \leq N\} \) are arbitrary linear,
CPTP maps with non-commuting outputs. In this case
it is well-known that we have to take into account the
fact that the evolving quantum system, originally living
in \( \mathcal{H}_S \), could be a part of a larger system, living in the
tensor product space \( \mathcal{H}_S \otimes \mathcal{H}_S \). In fact, while in principle
the ancillary system \( S' \) can be arbitrary, without loss
of generality we can restrict it to be a copy of \( S \), i.e.,
\( \mathcal{H}_S \cong \mathcal{H}_S \). Accordingly, we reformulate Definition 1 in
order to allow for such possible extensions:

**Definition 2.** A given discrete-time quantum stochastic
process \( \mathcal{F} = \{N^i; 1 \leq i \leq N\} \) is said to be completely
information decreasing if and only if, for all finite en-
sembles \( \mathcal{E} = \{p(x); \rho^x_{S,S'}\}_x \) of states on \( \mathcal{H}_S \otimes \mathcal{H}_S \),
the sequence of guessing probabilities \( [P_{\text{guess}}(\mathcal{E}_i)]_{i \in \mathbb{N}} \),
where \( \mathcal{E}_i := \{p(x), (id_{S'} \otimes N^i)(\rho^x_{S,S'})\}_x \), is mono-
tonically non-increasing, i.e., \( P_{\text{guess}}(\mathcal{E}_i) \geq P_{\text{guess}}(\mathcal{E}_{i+1}) \) for
all \( 1 \leq i \leq N - 1 \).
Using the partial ordering notation $\succeq$ previously introduced, we can equivalently say that the process described by $\mathcal{F} = \{\mathcal{N}_i; 1 \leq i \leq N\}$ is completely information decreasing if and only if $\{\mathcal{N}_1 \otimes \mathcal{N}_2 \succeq \mathcal{N}_1 \otimes \mathcal{N}_2 \succeq \cdots \succeq \mathcal{N}_1 \otimes \mathcal{N}_2\}$. Then, the following statement holds (for a more general statement and its proof, see Lemma 3 in the Supplemental Material [33]):

**Proposition 3** (The quantum case). A given discrete-time quantum stochastic process is Markovian if and only if it is completely information decreasing.

**Discussion.**—Propositions 2 and 3 above establish that Markovianity of a discrete-time stochastic process is equivalent to a monotonic decrease of information. They hence provide an information-theoretic underpinning of the memoryless property which is the key feature of a Markov chain. This equivalence is also valid in the case of continuous-time stochastic processes, since the latter can be obtained from the discrete-time setting by considering instants in time which are arbitrarily close to each other.

In this respect, our approach can be seen as a generalization of an idea first proposed by Breuer, Laine, and Piilo in [2]. They characterized stochastic processes by tracking the change in the distinguishability of two different initial states of the system under dynamic evolution. However, while in [2] only equiprobable pairs of states were considered, here we track the evolution of arbitrary ensembles of quantum states, i.e., ensembles consisting of more than two states in general, with arbitrary apriori probabilities and possibly living on a bipartite Hilbert space. This is the reason why our condition is strong enough to be equivalent to Markovianity, while the criterion proposed in [2] is only necessary but not sufficient, as explicitly shown by Chruściński, Kossakowski, and Rivas in [9]. In fact, building upon the results of [15], Chruściński et al. also proposed a strengthened version of the criterion of Breuer et al.. However, their criterion is valid only for quantum stochastic processes which involve quantum channels that are all invertible (as linear maps). This rules out physically relevant situations like, for example, semiclassical processes. On the contrary, our approach does not require any assumption on the underlying stochastic process: the channels in $\mathcal{F}$ can be completely arbitrary.

Note that Proposition 3 also provides an operational characterization of reversible stochastic processes, as those for which the guessing probability is constant, i.e.,

$$P_{\text{guess}}(\tilde{E}_i) = P_{\text{guess}}(\tilde{E}_{i+1}), \quad 1 \leq i \leq N - 1,$$

for any initial ensemble. The above equality implies the existence of not only ‘direct propagators’, i.e., quantum channels $\mathcal{C}_i$ such that $\mathcal{C}_i \otimes \mathcal{N} \succeq \mathcal{N} \succeq \mathcal{N} \otimes \mathcal{C}_i$, but also the existence of ‘reverse propagators’, i.e., quantum channels $\overline{\mathcal{C}_i}$ such that $\overline{\mathcal{C}_i} \otimes \mathcal{N} = \mathcal{N} \otimes \overline{\mathcal{C}_i}$, with the convention that $\mathcal{N}^0 = \mathcal{id}$. In other words, a stochastic process which preserves information has to be reversible. Since the only CPTP maps which are reversible are unitary ones (see, e.g., Ref. [7]), we arrive at the following conclusion: the only stochastic processes which preserve information perfectly are those describing the evolution of closed systems. This is in keeping with intuition since a closed system has no environment into which information can leak, and whose evolution is, therefore, automatically Markovian.

A further observation is that, according to the results in [4, 5], the ensembles of bipartite states $\mathcal{E}$ used in Definition 3 can without loss of generality be restricted to ensembles of separable states. In principle this may simplify the experimental assessment of Markovianity, since entanglement is not needed. This is because, as shown in Ref. [5], the identity channel $\mathcal{id}_{\mathcal{S}}$, used in Definition 3 to define the partial ordering relation $\{\mathcal{id}_S \otimes \mathcal{N}_i \succeq \mathcal{id}_S \otimes \mathcal{N}_{i+1}\}$, can be replaced, without loss of generality, with some other noisy channel $\mathcal{M}_S : \mathcal{L}(\mathcal{H}_S) \to \mathcal{L}(\mathcal{H}_S)$, under the sole condition that $\mathcal{M}_S$ is complete [3], i.e., its image spans the whole $\mathcal{L}(\mathcal{H}_S)$. Since there exist complete quantum channels which are entanglement-breaking (e.g., a depolarizing channel $\mathcal{D}(\omega) = \omega + (1 - \epsilon)\mathcal{I}$ with sufficiently small but nonzero $\epsilon$), we arrive at the conclusion that in Definition 3 it actually suffices to consider bipartite states $\mathcal{P}_{\mathcal{S}_0}$ which are separable.

**Acknowledgements**—The authors acknowledge helpful discussions with David Reeb and Sergii Strelchuk. F.B. acknowledges the Statistical Laboratory of the University of Cambridge where this research was done.

*buscemi@is.nagoya-u.ac.jp
n.datta@statslab.cam.ac.uk

[1] D. Blackwell, Ann. Math. Stat. 24, 265 (1953).
[2] H.-P. Breuer, E.-M. Laine and J. Piilo, Phys. Rev. Lett. 103, 210401 (2009).
[3] F. Buscemi, Phys. Rev. Lett. 108, 200401 (2012).
[4] F. Buscemi, Comm. Math. Phys. 310, 625 (2012).
[5] F. Buscemi, On complete positivity, Markovianity and the quantum data-processing inequality, in the presence of initial system-environment correlations. arXiv:1307.0363 [quant-ph].
[6] F. Buscemi, N. Datta and S. Strelchuk, Game-theoretic characterization of antidegradable channels. arXiv:1404.0277 [quant-ph].
[7] F. Buscemi, M. Keyl, G.M. D’Ariano, P. Perinotti and R.F. Werner, J. Math. Phys. 46, 82109 (2005).
[8] A. Cheles, The Quantum Blackwell Theorem and Minimum Error State Discrimination. arXiv:0907.0866 [quant-ph].
[9] D. Chruściński, A. Kossakowski and À. Rivas, Phys. Rev. A 83, 052128 (2011).
[10] T.M. Cover and J.A. Thomas, Elements of information theory (John Wiley & Sons, 2012).
[11] I. Csizsár, Per. Math. Hung. 2, 191-213 (1972).
[12] A. D’Arrigo, R. Lo Franco, G. Benenti, E. Paladino and
G. Falci, Ann. Phys. 350, 211 (2014).

[13] M.J.W. Hall, J.D. Cresser, L. Li and E. Andersson, Phys. Rev. A, 89, 042120 (2014).

[14] A.S. Holevo, J. Multivar. Anal. 3, 337 (1973).

[15] A Kossakowski, Bull. Acad. Sci. Math. 20, 1021 (1972).

[16] K. Kraus, Annals of Phys. 64, 311 (1971).

[17] B.-H. Liu, L. Li, Y.-F. Huang, C.-F. Li, G.-C. Guo, E.-M. Laine, H.-P. Breuer and J. Piilo, Nature Phys. 7, 931 (2011).

[18] M.A. Nielsen and I.L. Chuang, Quantum computation and quantum information. (Cambridge University Press, 2010).

[19] J.R. Norris, Markov chains. (Cambridge University Press, 1998).

[20] E. Parzen, Stochastic Processes. (SIAM, 1999).

[21] F. Petruccione and H.-P. Breuer, The theory of open quantum systems. (Oxford University Press, 2002.)

[22] A. Rivas, S.F. Huelga and M.B. Plenio, Rep. Prog. Phys. 77, 094001 (2014).

[23] Á. Rivas, S.F. Huelga and M.B. Plenio, Phys. Rev. Lett 105, 050403 (2010).

[24] E. Shmaya, J. Phys. A: Math. Gen. 38, 9717 (2005).

[25] A. Smirne, A. Stabile and B. Vacchini, Physica Scripta T153, 014057 (2013).

[26] W.F. Stinespring, Proc. Am. Math. Soc. 6, 211 (1955).

[27] E. Torgersen, Comparison of statistical experiments. (Cambridge University Press, 1991).

[28] E.N. Torgersen, Prob. Th. and Related Fields 16, 219 (1970).

[29] A.R. Usha Devi, A.K. Rajagopal and Sudha, Phys. Rev. A 83, 022109 (2011).

[30] B. Vacchini, A. Smirne, E.-M. Laine, J. Piilo and H.-P. Breuer, New J. Phys 13, 093004 (2011).

[31] J.-S. Xu, K. Sun, C.-F. Li, X.-Y. Xu, G.-C. Guo, E. Andersson, R. Lo Franco and G. Compagno, Nat. Comm. 4, 2851 (2013).

[32] H.P. Yuen, R.S. Kennedy and M.Lax, IEEE Trans. Inf. Th. 21, 125 (1975).

[33] See the Supplemental Material.
The equivalence between Markovianity and monotonic decrease of information. Supplemental Material. — 1/2

The results in this section are based on Refs [3][1][6][8][24]. In what follows, we adopt a more abstract viewpoint and consider the case, in which channels (CPTP maps) can have different initial and final spaces. A part from this, the meaning of symbols and notations remains unchanged. Another difference with the main text is that, here, we consider only two channels at a time, rather than a sequence of them. Notice, however, that this does not cause any loss in generality, since Markovianity is, rather than a global property of the discrete-time stochastic process, a property involving only one channel and its immediate successor in the sequence.

**Lemma 1 (3).** For any pair of CPTP maps \( \mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B) \) and \( \mathcal{N}' : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{B'}) \), the following are equivalent:

1. for any alphabet \( \mathcal{X} = \{x\} \) and for any ensemble \( \mathcal{E} = \{p(x), \rho_A^x\}_{x \in \mathcal{X}} \) of density matrices in \( \mathcal{D}(\mathcal{H}_A) \),
   \[
P_{\text{guess}}[\mathcal{N}(\mathcal{E})] \geq P_{\text{guess}}[\mathcal{N}'(\mathcal{E})],
   \] (S.1)
   where \( \mathcal{N}(\mathcal{E}) \) is a shorthand notation for the ensemble \( \{p(x), \mathcal{N}(\rho_A^x)\}_{x \in \mathcal{X}} \), and analogously \( \mathcal{N}'(\mathcal{E}) \);

2. for any alphabet \( \mathcal{Y} = \{y\} \) and for any POVM \( \{Q_B^y\}_{y \in \mathcal{Y}} \), there exists a POVM \( \{P_B^y\}_{y \in \mathcal{Y}} \) such that
   \[
   \text{Tr}[\mathcal{N}'(\rho_A) \ Q_B^y] = \text{Tr}[\mathcal{N}(\rho_A) \ P_B^y],
   \] (S.2)
   for all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \) and for all \( y \in \mathcal{Y} \).

Whenever the above conditions hold, we say that \( \mathcal{N} \) is more informative than \( \mathcal{N}' \), denoted by \( \mathcal{N} \preceq \mathcal{N}' \).

**Proof.** The above result is a direct consequence of Theorem 3 of Ref. [3]: point number 1 above corresponds to point number 4 there – point number 2 above corresponds to point number 3 there.

**Lemma 2.** Let \( \mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B) \) and \( \mathcal{N}' : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{B'}) \) be two CPTP maps. Suppose that the output of \( \mathcal{N}' \) is abelian, i.e., \( [\mathcal{N}'(\rho), \mathcal{N}'(\sigma)] = 0 \), for any \( \rho, \sigma \in \mathcal{D}(\mathcal{H}_A) \).

Then, \( \mathcal{N} \preceq \mathcal{N}' \) if and only if there exists a third CPTP map \( \mathcal{C} : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_{B'}) \) such that
\[
\mathcal{N}' = \mathcal{C} \circ \mathcal{N}.
\]

**Proof.** Here we prove only the ‘only if’ part of the statement, as the ‘if’ part is trivial. The proof is based on the analogous result for bipartite states derived in Ref. [3].

Since the outputs of \( \mathcal{N}' \) are all commuting, it is possible to find a basis \( \{\ket{i_{B'}} \in \mathcal{H}_{B'}\}_i \) that diagonalizes them all simultaneously. A simple identity then gives:
\[
\mathcal{N}'(\rho_A) = \sum_i \ket{i_{B'}}\bra{i_{B'}} \text{Tr}[\mathcal{N}'(\rho_A) \ket{i_{B'}}\bra{i_{B'}}],
\] (S.3)
for all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \). On the other hand, since \( \mathcal{N} \preceq \mathcal{N}' \), and since \( \{\ket{i_{B'}}\bra{i_{B'}}\}_i \) constitutes a well-defined POVM on \( \mathcal{H}_{B'} \), we know that there exists a POVM \( \{P_B^i\} \) on \( \mathcal{H}_B \) such that
\[
\text{Tr}[\mathcal{N}'(\rho_A) \ket{i_{B'}}\bra{i_{B'}}] = \text{Tr}[\mathcal{N}(\rho_A) \ P_B^i],
\]
for all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \) and all \( i \). Inserting the above equation into (S.3), one obtains the identity
\[
\mathcal{N}'(\rho_A) = \sum_i \ket{i_{B'}}\bra{i_{B'}} \text{Tr}[\mathcal{N}(\rho_A) \ P_B^i],
\]
valid for all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \), which can be equivalently written as \( \mathcal{N}' = \mathcal{C} \circ \mathcal{N} \) upon defining the CPTP map \( \mathcal{C} \) as
\[
\mathcal{C}(\bullet_B) := \sum_i \ket{i_{B'}}\bra{i_{B'}} \text{Tr}[\bullet_B \ P_B^i].
\]
The above equation shows, in particular, that the map \( \mathcal{C} : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_{B'}) \) is a CPTP map defined everywhere, as claimed in the statement.

**Lemma 3.** Given a pair of CPTP maps \( \mathcal{N} : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_B) \) and \( \mathcal{N}' : \mathcal{L}(\mathcal{H}_A) \to \mathcal{L}(\mathcal{H}_{B'}) \), let \( \mathcal{H}_{B''} \) be an auxiliary Hilbert space isomorphic with \( \mathcal{H}_{B'} \), i.e., \( \mathcal{H}_{B''} \cong \mathcal{H}_{B'} \), and \( \text{id} : \mathcal{L}(\mathcal{H}_{B''}) \to \mathcal{L}(\mathcal{H}_{B''}) \) the corresponding identity CPTP map.

Then, \( \text{id} \otimes \mathcal{N} \succeq \text{id} \otimes \mathcal{N}' \) if and only if there exists a third CPTP map \( \mathcal{C} : \mathcal{L}(\mathcal{H}_B) \to \mathcal{L}(\mathcal{H}_{B'}) \) such that
\[
\mathcal{N}' = \mathcal{C} \circ \mathcal{N}.
\]
Proof. Here we prove only the ‘only if’ part of the statement, as the ‘if’ part is trivial. The proof presented here is based on a series of results appeared in Ref. [3, 5, 6, 24].

By hypothesis, it holds that \( \text{id} \otimes \mathcal{N} \geq \text{id} \otimes \mathcal{N}' \), which implies, in particular, that, for any finite alphabet \( \mathcal{Y} = \{ y \} \) and any POVM \( \{ Q^y_{B'B'} \} \), there exists a POVM \( \{ P^y_{B'B'} \} \) such that

\[
\text{Tr} \left[ (\omega_{B''} \otimes \mathcal{N}'(\rho_A)) \ Q^y_{B'B'} \right] = \text{Tr} \left[ (\omega_{B''} \otimes \mathcal{N}(\rho_A)) \ P^y_{B'B'} \right],
\]

for all \( y \in \mathcal{Y} \), all \( \omega_{B''} \in \mathcal{D}(\mathcal{H}_{B''}) \), and all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \). Upon introducing another auxiliary Hilbert space \( \mathcal{H}_{B'''} \equiv \mathcal{H}_{B''} \) and a maximally entangled state \( |\Phi^+_{B''B'}\rangle \in \mathcal{H}_{B''} \otimes \mathcal{H}_{B''} \), the condition expressed in Eq. (S.4) can be rewritten as follows: for any alphabet \( \mathcal{Y} = \{ y \} \) and any POVM \( \{ Q^y_{B'B'} \} \), there exists a POVM \( \{ P^y_{B'B'} \} \) such that

\[
\text{Tr} \left[ |\Phi^+_{B''B'}\rangle \langle \Phi^+_{B''B'}| \otimes \mathcal{N}'(\rho_A) \right] \ {\Omega}_{B''} \otimes \mathcal{Q}^y_{B'B'}
\]

\[
= \text{Tr} \left[ |\Phi^+_{B''B'}\rangle \langle \Phi^+_{B''B'}| \otimes \mathcal{N}(\rho_A) \right] \ {\Omega}_{B''} \otimes \mathcal{P}^y_{B'B'} \],
\]

for all \( y \in \mathcal{Y} \), all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \), and all \( 0 \leq \Omega_{B''} \in \mathcal{L}(\mathcal{H}_{B''}) \).

We now make use of the simple fact that, \( \text{Tr}[XA] = \text{Tr}[YA] \) for all \( A \geq 0 \) if and only if \( X = Y \), to reformulate condition (S.5), involving positive numbers, into a condition involving operators: for any alphabet \( \mathcal{Y} = \{ y \} \) and any POVM \( \{ Q^y_{B'B'} \} \), there exists a POVM \( \{ P^y_{B'B'} \} \) such that

\[
\text{Tr}_{B'B'} \left[ |\Phi^+_{B''B'}\rangle \langle \Phi^+_{B''B'}| \otimes \mathcal{N}'(\rho_A) \right] \ {\Omega}_{B''} \otimes \mathcal{Q}^y_{B'B'}
\]

\[
= \text{Tr}_{B'B'} \left[ |\Phi^+_{B''B'}\rangle \langle \Phi^+_{B''B'}| \otimes \mathcal{N}(\rho_A) \right] \ {\Omega}_{B''} \otimes \mathcal{P}^y_{B'B'} \],
\]

for all \( y \in \mathcal{Y} \) and all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \).

Now we recall the protocol of (generalized) teleportation of Ref. [S1], according to which one can always choose the alphabet \( \mathcal{Y} = \{ y \} \) and the POVM \( \{ Q^y_{B'B'} \} \) in Eq. (S.6) such that

\[
\mathcal{N}'(\rho_A) = \sum_y U^y_B \circ \text{Tr}_{B'B'} \left[ |\Phi^+_{B''B'}\rangle \langle \Phi^+_{B''B'}| \otimes \mathcal{N}'(\rho_A) \right] \ {\Omega}_{B''} \otimes \mathcal{Q}^y_{B'B'}
\]

for all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \), where the maps \( U^y_B : \mathcal{L}(\mathcal{H}_{B''}) \rightarrow \mathcal{L}(\mathcal{H}_B) \) are suitable unitary CPTP maps, i.e., \( U^y(\bullet) = U^y_\dagger U^y \) with \( U^y_\dagger U^y = 1_{B''} \). Then, condition (S.6) guarantees the existence of a POVM \( \{ P^y_{B'B'} \} \) such that

\[
\mathcal{N}'(\rho_A) = \sum_y U^y_B \circ \text{Tr}_{B'B'} \left[ |\Phi^+_{B''B'}\rangle \langle \Phi^+_{B''B'}| \otimes \mathcal{N}(\rho_A) \right] \ {\Omega}_{B''} \otimes \mathcal{P}^y_{B'B'}
\]

for all \( \rho_A \in \mathcal{D}(\mathcal{H}_A) \). The above identity can be equivalently written as the channel identity \( \mathcal{N}' = \mathcal{C} \circ \mathcal{N} \), upon introducing the map \( \mathcal{C} : \mathcal{L}(\mathcal{H}_B) \rightarrow \mathcal{L}(\mathcal{H}_B) \) defined as

\[
\mathcal{C} \left( \bullet_B \right) := \sum_y U^y_B \circ \text{Tr}_{B'B'} \left[ |\Phi^+_{B''B'}\rangle \langle \Phi^+_{B''B'}| \otimes \bullet_B \right] \ {\Omega}_{B''} \otimes \mathcal{P}^y_{B'B'} \].

The above equation shows, in particular, that the map \( \mathcal{C} \) is a CPTP map defined everywhere, as claimed in the statement.

Extended bibliography

[S1] S.L. Braunstein, G.M. D’Ariano, G.J. Milburn and M.F. Sacchi, Phys. Rev. Lett., 84, 3486 (2000).