ISSN 0032-9460, Problems of Information Transmission, 2016, Vol. 52, No. 3, pp. 201–213. © Pleiades Publishing, Inc., 2016.
Original Russian Text © F. Buscemi, 2016, published in Problemy Peredachi Informatsii, 2016, Vol. 52, No. 3, pp. 3–16.

INFORMATION THEORY

Degradable Channels, Less Noisy Channels, and Quantum Statistical Morphisms: An Equivalence Relation[1]

F. Buscemi

Department of Computer Science and Mathematical Informatics, Nagoya University, Nagoya, Japan
e-mail: buscemi@is.nagoya-u.ac.jp

Received February 9, 2016

Abstract—Two partial orderings among communication channels, namely "being degradable into" and "being less noisy than," are reconsidered in the light of recent results about statistical comparisons of quantum channels. Though our analysis covers at once both classical and quantum channels, we also provide a separate treatment of classical noisy channels and show how in this case an alternative self-contained proof can be constructed, with its own particular merits with respect to the general result.

DOI: 10.1134/S0032946016030017

1. INTRODUCTION

Given two channels, it is natural to ask which one is "better." As one soon realizes, ordering channels according to their capacity would be too limited in scope, since there may be other measures of "goodness" that are more relevant than capacity for the task at hand. This is even more so for *quantum* channels, for which many inequivalent capacities exist [1]. Indeed, it does not require too much imagination to come up with uncountably many such "comparisons," and each comparison will only induce a *partial*, rather than a *total*, ordering between channels. This fact should not come as a worrying surprise: channels are highly dimensional objects, and any total ordering can only be extremely coarse—often too coarse to be of any use in practice. Nonetheless, it is true that some comparisons are more natural, more compelling, or just mathematically simpler than others, so that they received more attention in the literature. This paper actually deals with two much studied comparisons, namely the partial orderings "being degradable into" and "being less noisy than," introduced in Definitions 1 and 2 below (for a compendium of many comparisons among discrete noisy channels, see [2]).

The goal of this work is to exhibit a connection between degradable channels and less noisy channels, beyond the obvious one "degradable implies less noisy." More explicitly, we show how a formal (but not substantial) modification in the definition of less noisy channels is sufficient to make the two orderings equivalent. Our result is proved in a general scenario, where channels are modeled as CPTP maps between operator algebras, thus covering quantum channels, classical channels (when input and output are commutative algebras), and also hybrid classical-to-quantum and quantum-to-classical channels.

Central to our approach is the notion of *quantum statistical morphisms*, i.e., linear maps between operator algebras that generalize in a statistical sense the idea of "post-processing" or "coarse-graining" (see Definition 5). The use of statistical morphisms allows us to prove our results under very mild assumptions, so that the quantum and classical cases are recovered as special cases of

[1] Supported in part by the JSPS KAKENHI, grant no. 26247016.

a single unifying framework. Such a top-down approach, besides being mathematically simpler, has the merit to clearly separate statistics from physics: indeed, it can immediately be applied to general probabilistic theories, since it does not rely on any particular feature of quantum theory like, for example, complete positivity.

The paper is organized as follows. In Section 2 we introduce the notation and some basic definitions. In Section 3 we introduce the concept of statistical morphisms and prove the fundamental equivalence relation. In Section 4 we specialize to the case of semiclassical channels. The case of discrete noisy classical channels is treated separately in Section 4.1, in a way that does not rely on any knowledge of the quantum case and allows the treatment of the approximate case, studied in the Appendix. In Section 5 we consider the case of fully quantum channels. Section 6 concludes the paper with some comments and possible future developments.

2. NOTATION AND DEFINITIONS

In what follows, all sets are finite and Hilbert spaces are finite-dimensional.

- Sets are denoted by $\mathcal{X}=\{x: x\in\mathcal{X}\}$, $\mathcal{Y}=\{y: y\in\mathcal{Y}\}$, etc.
- A probability distribution over $\mathcal{X}$ is a function $p\colon \mathcal{X}\to[0,1]$ such that $\sum_x p(x)=1$.
- The *set of all probability distributions* over $\mathcal{X}$ is denoted by $\mathrm{P}(\mathcal{X})$.
- Random variables are labeled by upper case letters X, Y, etc. with ranges $\mathcal{X}=\{x\}$, $\mathcal{Y}=\{y\}$, etc.
- Discrete noisy channels are identified with the associated conditional probability distributions.
- Quantum systems are labeled by upper case letters Q, R, etc., and the associated Hilbert spaces are denoted by $\mathcal{H}_Q$, $\mathcal{H}_R$, etc. Dimensions are denoted as $d_Q \stackrel{\text{def}}{=} \dim\mathcal{H}_Q$ etc.
- The *set of linear operators* acting on a Hilbert space $\mathcal{H}$ is denoted by $\mathrm{L}(\mathcal{H})$. The *identity operator* is denoted by $\mathbb{1}$.
- States of Q are represented by *density operators*, i.e., operators $\rho\in\mathrm{L}(\mathcal{H})$ such that $\rho\ge 0$ and $\mathrm{Tr}[\rho]=1$.
- The *set of density operators* acting on a Hilbert space $\mathcal{H}$ is denoted by $\mathrm{S}(\mathcal{H})$.
- A positive operator-valued measure (*POVM*) is a function $P\colon\mathcal{X}\to\mathrm{L}(\mathcal{H})$ such that $P(x)\ge 0$ and $\sum_x P(x)=\mathbb{1}$. For the sake of readability, we will often write the argument x as a superscript, i.e., P^x rather than $P(x)$.
- The *set of POVMs* from $\mathcal{X}$ to $\mathrm{L}(\mathcal{H})$ is denoted by $\mathrm{M}(\mathcal{X},\mathcal{H})$.
- *Quantum channels* are completely positive trace-preserving (CPTP) linear maps $\mathcal{N}\colon\mathrm{L}(\mathcal{H}_Q)\to\mathrm{L}(\mathcal{H}_R)$. The *range* of a channel $\mathcal{N}$ is defined as the image of $\mathrm{L}(\mathcal{H}_Q)$ under the action of $\mathcal{N}$, namely, the set $\{\mathcal{N}(X): X\in\mathrm{L}(\mathcal{H}_Q)\}$. The *identity map* is denoted by id.
- The *set of quantum channels* from $\mathrm{L}(\mathcal{H}_Q)$ to $\mathrm{L}(\mathcal{H}_R)$ is denoted by $\mathrm{C}(\mathcal{H}_Q,\mathcal{H}_R)$.
- Given a linear map $\mathcal{L}\colon\mathrm{L}(\mathcal{H}_Q)\to\mathrm{L}(\mathcal{H}_R)$, its *trace dual* is the linear map $\mathcal{L}^*\colon\mathrm{L}(\mathcal{H}_R)\to\mathrm{L}(\mathcal{H}_Q)$ defined by the relation
$$\mathrm{Tr}[\mathcal{L}^*(X_R)\,Y_Q]\stackrel{\text{def}}{=}\mathrm{Tr}[X_R\,\mathcal{L}(Y_Q)],$$
for all $Y_Q\in\mathrm{L}(\mathcal{H}_Q)$ and all $X_R\in\mathrm{L}(\mathcal{H}_R)$. Then $\mathcal{N}$ is trace-preserving if and only if $\mathcal{N}^*$ is *unit-preserving*, i.e., $\mathcal{N}^*(\mathbb{1}_R)=\mathbb{1}_Q$. Moreover, we say that a linear map $\mathcal{L}$ is *Hermitian* if and only if, for any $X=X^\dagger$, $\mathcal{L}(X)=\mathcal{L}(X)^\dagger$.
- A *classical-to-quantum (cq) channel* is a function $\mathcal{E}\colon\mathcal{X}\to\mathrm{S}(\mathcal{H})$. We will usually denote the density operators $\mathcal{E}(x)$ by ρ^x, σ^x, etc. Equivalently, a cq-channel $\mathcal{E}$ will be denoted as a family of density operators $\mathcal{E}=\{\rho^x: x\in\mathcal{X}\}$.
- A *classical-quantum (cq) state* is a bipartite density operator describing a quantum system Q correlated with a random variable X. Since random variables can be seen as commuting density operators, we will represent cq-states as, e.g., $\rho_{XQ}=\sum_{x\in\mathcal{X}}p(x)|x\rangle\langle x|_X\otimes\rho_Q^x$, where the unit vectors $\{|x\rangle: x\in\mathcal{X}\}$ are all orthogonal.

- For a given bipartite density operator $\rho_{RQ} \in \mathrm{S}(\mathcal{H}_R \otimes \mathcal{H}_Q)$, its *conditional min-entropy* is defined as
$$H_{\min}(R\,|\,Q)_\rho \overset{\text{def}}{=} -\inf_{\sigma_Q \in \mathrm{S}(\mathcal{H}_Q)} \inf\{\lambda \in \mathbb{R} :\ \rho_{RQ} \le 2^\lambda \mathbb{1}_R \otimes \sigma_Q\}.$$
We will use in particular the fact that [3]
$$2^{-H_{\min}(R|Q)_\rho} = d_R \max_{\mathcal{N} \in \mathrm{C}(\mathcal{H}_Q, \mathcal{H}_{R'})} F^2\big((\mathrm{id}_R \otimes \mathcal{N})\rho_{RQ}, \Phi^+_{RR'}\big),$$
where $\mathcal{H}_{R'} \cong \mathcal{H}_R$, $F^2(\rho,\sigma) \overset{\text{def}}{=} \|\sqrt{\rho}\sqrt{\sigma}\|_1^2$, and $\Phi^+_{RR'} \overset{\text{def}}{=} d_R^{-1} \sum_{i,j=1}^{d_R} |i_R\rangle|i_{R'}\rangle\langle j_R|\langle j_{R'}|$, for some orthonormal basis $\{|i\rangle\}$ of $\mathcal{H}_R$. In the case of a cq-state $\rho_{XQ} = \sum_{x\in\mathcal{X}} p(x)|x\rangle\langle x|_X \otimes \rho_Q^x$, the above formula becomes equivalent to
$$2^{-H_{\min}(X|Q)_\rho} = \max_{P \in \mathrm{M}(\mathcal{X}, \mathcal{H}_Q)} \sum_{x\in\mathcal{X}} p(x)\,\mathrm{Tr}\big[\rho_Q^x\, P_Q^x\big] \overset{\text{def}}{=} P_{\text{guess}}(X\,|\,Q)_\rho,$$
namely, the expected *guessing probability*, i.e., the probability of correctly guessing the value of X having access only to the quantum system Q.

2.1. "Degradable" and "Less Noisy" Channels

In the classical case, the following definitions can be found in [4–7].

Definition 1 (degradable channels). Given two discrete channels $p(y\,|\,x)$ and $p'(z\,|\,x)$, p is said to be *degradable* into p' whenever there exists a discrete channel $q(z\,|\,y)$ such that
$$p'(z\,|\,x) = \sum_{y\in\mathcal{Y}} q(z\,|\,y)p(y\,|\,x).$$

Definition 2 (less noisy channels). Given two discrete channels $p(y\,|\,x)$ and $p'(z\,|\,x)$, p is said to be *less noisy* than p' whenever, for any discrete random variable U, any probability distribution $q(u)$, and any channel $q(x\,|\,u)$, the joint input-output probability distributions $q(u)q(x\,|\,u)p(y\,|\,x)$ and $q(u)q(x\,|\,u)p'(z\,|\,x)$ satisfy
$$H(U\,|\,Y) \le H(U\,|\,Z).$$

If p_1 is degradable into p_2, then p_1 is less noisy than p_2: the proof is a simple consequence of the data-processing inequality. Counterexamples are known for the converse [5], namely, one channel can be less noisy than another without being degradable. A consequence of the results presented here is that it is sufficient to replace H with $H_{\min}$ in Definition 2 in order to make the ordering "less noisy" (now defined with respect to $H_{\min}$) equivalent to the ordering "degradable." This fact is formalized in Corollary 3 below. (The reader interested in the classical case only can directly skip to Section 4.1: there, Corollary 3 is provided with an independent, self-contained proof, which does not rely on any idea developed for the general noncommutative case. Moreover, such a proof allows the treatment of the approximate case, which is studied in the Appendix.)

In fact, our analysis will not be limited to the case of classical noisy channels but will include some results valid for quantum channels too. We hence generalize Definitions 1 and 2 to the quantum case as follows (but compare with [8]).

Definition 3 (degradable quantum channels). Given two CPTP maps $\mathcal{N}\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_R)$ and $\mathcal{N}'\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_S)$, $\mathcal{N}$ is said to be *degradable* into $\mathcal{N}'$ whenever there exists a CPTP map $\mathcal{T}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ such that
$$\mathcal{N}' = \mathcal{T} \circ \mathcal{N}.$$

Definition 4 (less noisy quantum channels). Given two CPTP maps $\mathcal{N}\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_R)$ and $\mathcal{N}'\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_S)$, $\mathcal{N}$ is said to be *less noisy* than $\mathcal{N}'$ whenever, for any discrete random variable U, any probability distribution $q(u)$, and any cq-channel $\mathcal{E} = \{\rho_Q^u : u \in \mathcal{U}\}$, the corresponding input-output cq-states

$$\sigma_{UR} \stackrel{\text{def}}{=} \sum_u q(u)|u\rangle\langle u|_U \otimes \mathcal{N}_Q(\rho_Q^u) \qquad \text{and} \qquad \tau_{US} \stackrel{\text{def}}{=} \sum_u q(u)|u\rangle\langle u|_U \otimes \mathcal{N}'_Q(\rho_Q^u)$$

satisfy

$$H(U\,|\,R)_\sigma \le H(U\,|\,S)_\tau.$$

This paper studies the relations between the notions of degradable channels and less noisy channels, both in classical and quantum information theory. In what follows we will show, in particular, how Definitions 2 and 4 can formally be modified so that the two partial orderings become equivalent. The results presented here are based on recent formulations of the Blackwell–Sherman–Stein theorem [9–11] for quantum systems [12–17].

3. STATISTICAL MORPHISMS AND A FUNDAMENTAL EQUIVALENCE RELATION

We begin with a definition, generalizing that given in [15].

Definition 5 (quantum statistical morphisms). Given a CPTP map $\mathcal{N}\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_R)$, a *statistical morphism* of $\mathcal{N}$ is a linear map $\mathcal{L}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ such that, for any finite outcome set $\mathcal{X}$ and any POVM $\{\bar{P}_S^x : x \in \mathcal{X}\}$, there exists another POVM $\{P_R^x : x \in \mathcal{X}\}$ such that

$$\mathrm{Tr}[(\mathcal{L}\circ\mathcal{N})(\rho_Q)\,\bar{P}_S^x] = \mathrm{Tr}[\mathcal{N}(\rho_Q)\,P_R^x], \quad \forall x \in \mathcal{X}, \quad \forall \rho_Q \in \mathrm{S}(\mathcal{H}_Q). \tag{1}$$

Remark 1. Clearly, a positive trace-preserving linear map is always a well-defined statistical morphism, *for any* channel. However, a map can be a statistical morphism *of some channel* without being positive and trace-preserving—in fact, statistical morphisms cannot even *be extended*, in general, to positive trace-preserving maps, as is shown in [18] by an explicit counterexample[2]. We can only say that, if $\mathcal{L}$ is a statistical morphism of $\mathcal{N}$, then $\mathcal{L}$ is positive and trace-preserving on the range of $\mathcal{N}$, namely, $\mathrm{Tr}[(\mathcal{L}\circ\mathcal{N})(X_Q)] = \mathrm{Tr}[\mathcal{N}(X_Q)]$, for all $X_Q \in \mathrm{L}(\mathcal{H}_Q)$, and, whenever $\mathcal{N}(X_Q) \ge 0$, $\mathrm{Tr}[(\mathcal{L}\circ\mathcal{N})(X_Q)\,\bar{P}_S] \ge 0$, for all $\bar{P}_S \ge 0$. The question then arises: Is any linear map $\mathcal{L}$ which is positive and trace-preserving on the range of a channel $\mathcal{N}$ a statistical morphism of $\mathcal{N}$? Again, the answer is negative. This is because, in order to guarantee that $\mathcal{L}$ is positive and trace-preserving on the range of $\mathcal{N}$, it would be sufficient to have (1) hold for binary POVMs (i.e., *effects*) $\{\bar{P}, \mathbb{1} - \bar{P}\}$ only, but this condition is known to be strictly weaker than that required in Definition 5, which must hold for any finite $\mathcal{X}$ [20]. The situation can thus be summarized as follows:

$$\text{PTP everywhere} \underset{\not\Leftarrow}{\Longrightarrow} \text{statistical morphism of } \mathcal{N} \underset{\not\Leftarrow}{\Longrightarrow} \text{PTP on range}(\mathcal{N}).$$

Remark 2. In what follows, when we say "trace-preserving statistical morphism," we mean a trace-preserving (everywhere) linear map that is, in particular, a statistical morphism (for some channel).

We are now ready to state a fundamental equivalence relation.

Proposition 1. *Given two CPTP maps $\mathcal{N}\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_R)$ and $\mathcal{N}'\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_S)$, the following are equivalent*:

[2] About this problem, see also [19].

(i) *For any discrete random variable* U*, any probability distribution* $q(u)$*, and any cq-channel* $\mathcal{E} = \{\rho_Q^u : u \in \mathcal{U}\}$*, the corresponding input-output cq-states*

$$\sigma_{UR} = \sum_{u\in\mathcal{U}} q(u)|u\rangle\langle u|_U \otimes \mathcal{N}(\rho_Q^u) \qquad \text{and} \qquad \tau_{US} = \sum_{u\in\mathcal{U}} q(u)|u\rangle\langle u|_U \otimes \mathcal{N}'(\rho_Q^u)$$

satisfy

$$H_{\min}(U\,|\,R)_\sigma \le H_{\min}(U\,|\,S)_\tau,$$

i.e., $P_{\text{guess}}(U\,|\,R)_\sigma \ge P_{\text{guess}}(U\,|\,S)_\tau$;

(ii) *There exists a Hermitian trace-preserving statistical morphism* $\mathcal{L}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ *of* $\mathcal{N}$ *such that*

$$\mathcal{N}' = \mathcal{L} \circ \mathcal{N}.$$

Note that point (i) in Proposition 1 looks exactly as the definition of less noisy channels (Definition 4), the only difference being the use of $H_{\min}$ in the place of H.

Proof. If point (ii) holds, then

$$\begin{aligned} P_{\text{guess}}(U\,|\,S)_\tau &= \max_{\bar{P}\in\mathrm{M}(\mathcal{U},\mathcal{H}_S)} \sum_{u\in\mathcal{U}} q(u)\,\mathrm{Tr}[\mathcal{N}'(\rho_Q^u)\,\bar{P}_S^u] \\ &= \max_{\bar{P}\in\mathrm{M}(\mathcal{U},\mathcal{H}_S)} \sum_{u\in\mathcal{U}} p(u)\,\mathrm{Tr}[(\mathcal{L}\circ\mathcal{N})(\rho_Q^u)\,\bar{P}_S^u] \\ &\le \max_{P\in\mathrm{M}(\mathcal{U},\mathcal{H}_R)} \sum_{u\in\mathcal{U}} p(u)\,\mathrm{Tr}[\mathcal{N}(\rho_Q^u)\,P_R^u] = P_{\text{guess}}(U\,|\,R)_\sigma, \end{aligned}$$

where the inequality is a consequence of Definition 5 above.

Conversely, assume that point (i) holds. As is already shown in [15–17], this implies that, for any POVM $\{\bar{P}_S^x : x \in \mathcal{X}\}$ on $\mathcal{H}_S$, there exists a POVM $\{P_R^x : x \in \mathcal{X}\}$ on $\mathcal{H}_R$ such that

$$\mathrm{Tr}[\mathcal{N}'(\rho_Q)\,\bar{P}_S^x] = \mathrm{Tr}[\mathcal{N}(\rho_Q)\,P_R^x], \tag{2}$$

for all $x \in \mathcal{X}$ and all $\rho_Q \in \mathrm{S}(\mathcal{H}_Q)$. Let us choose $\{\bar{P}_S^x : x \in \mathcal{X}\} \in \mathrm{M}(\mathcal{X},\mathcal{H}_S)$ to be an informationally complete POVM, i.e., such that $\mathrm{span}\{\bar{P}_S^x : x \in \mathcal{X}\} = \mathrm{L}(\mathcal{H}_S)$. Let $\{P_R^x : x \in \mathcal{X}\} \in \mathrm{M}(\mathcal{X},\mathcal{H}_R)$ be the corresponding POVM, satisfying (2), and define a linear map $\mathcal{L}^*\colon \mathrm{L}(\mathcal{H}_S) \to \mathrm{L}(\mathcal{H}_R)$ by

$$\mathcal{L}^*(\bar{P}_S^x) \stackrel{\text{def}}{=} P_R^x, \quad x \in \mathcal{X}.$$

Such a map is uniquely defined, since $\{\bar{P}_S^x : x \in \mathcal{X}\}$ is a basis for $\mathrm{L}(\mathcal{H}_S)$, and it is unit-preserving by the construction, implying that its trace dual $\mathcal{L}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ is trace-preserving. In order to prove that $\mathcal{L}$ is also Hermitian, let $\{\Theta_S^x : x \in \mathcal{X}\}$ be the set of Hermitian operators in $\mathrm{L}(\mathcal{H}_S)$ such that

$$X_S = \sum_{x\in\mathcal{X}} \mathrm{Tr}[X_S\,\bar{P}_S^x]\Theta_S^x,$$

for all $X_S \in \mathrm{L}(\mathcal{H}_S)$. This implies that, for any $Y_R = Y_R^\dagger$ in $\mathrm{L}(\mathcal{H}_R)$,

$$\mathcal{L}(Y_R) = \sum_{x\in\mathcal{X}} \mathrm{Tr}[\mathcal{L}(Y_R)\,\bar{P}_S^x]\Theta_S^x = \sum_{x\in\mathcal{X}} \mathrm{Tr}[Y_R\,\mathcal{L}^*(\bar{P}_S^x)]\Theta_S^x = \sum_{x\in\mathcal{X}} \mathrm{Tr}[Y_R\,P_R^x]\Theta_S^x = \sum_{x\in\mathcal{X}} \lambda_x\Theta_S^x$$

with $\lambda_x \in \mathbb{R}$; i.e., $\mathcal{L}(Y_R)$ is Hermitian too. Hence, $\mathcal{L}$, as defined above, is a Hermitian trace-preserving linear map. We only need to show that $\mathcal{N}' = \mathcal{L} \circ \mathcal{N}$ and that $\mathcal{L}$ is a well-defined statistical morphism of $\mathcal{N}$.

In order to show that $\mathcal{N}' = \mathcal{L} \circ \mathcal{N}$, we notice that the condition expressed in (2) can be reformulated as follows: for any $\rho_Q \in \mathrm{S}(\mathcal{H}_Q)$ and any $x \in \mathcal{X}$,

$$\mathrm{Tr}[\mathcal{N}'(\rho_Q)\,\bar{P}_S^x] = \mathrm{Tr}[\mathcal{N}(\rho_Q)\,P_R^x] = \mathrm{Tr}[\mathcal{N}(\rho_Q)\,\mathcal{L}^*(\bar{P}_S^x)] = \mathrm{Tr}[(\mathcal{L} \circ \mathcal{N})(\rho_Q)\,\bar{P}_S^x].$$

Since $\{\bar{P}_S^x : x \in \mathcal{X}\}$ in the above equation is informationally complete, we have that $\mathcal{N}'(\rho_Q) = (\mathcal{L} \circ \mathcal{N})(\rho_Q)$, for all ρ_Q, i.e., $\mathcal{N}' = \mathcal{L} \circ \mathcal{N}$. Thus we also know that the condition expressed in (2) above automatically implies that $\mathcal{L}$ is a well-defined statistical morphism. $\triangle$

In other words, Proposition 1 states that replacing H with $H_{\min}$ in Definition 4 is sufficient to conclude a *weaker form of degradability*, in the sense that the degrading map is not a quantum channel but a Hermitian trace-preserving statistical morphism. In what follows, we will see when one can conclude that the degrading map is in fact CPTP.

Before proceeding, however, we specialize Proposition 1 to the case of cq-channels, which can always be seen as CPTP maps on commuting input subalgebras. We start by simplifying the definition of a statistical morphism as follows (this was the original definition given in [15]).

Definition 6. Given a cq-channel $\mathcal{E}\colon \mathcal{X} \to \mathrm{S}(\mathcal{H}_R)$ with $\mathcal{E} = \{\sigma_R^x : x \in \mathcal{X}\}$, a *statistical morphism* of $\mathcal{E}$ is a linear map $\mathcal{L}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ such that, for any $\mathcal{Y}$ and any POVM $\{\bar{P}_S^y : y \in \mathcal{Y}\}$, there exists a corresponding POVM $\{P_R^y : y \in \mathcal{Y}\}$ such that $\mathrm{Tr}[\mathcal{L}(\sigma_R^x)\,\bar{P}_S^y] = \mathrm{Tr}[\sigma_R^x\,P_R^y]$.

Remark 3. Note that, in order for $\mathcal{L}$ to be a well-defined statistical morphism of $\mathcal{E}$, it is not sufficient that $\mathcal{L}(\sigma_R^x) \in \mathrm{S}(\mathcal{H}_S)$ for all $x \in \mathcal{X}$. In particular, such a map is not, in general, positive on the whole span$\{\sigma_R^x : x \in \mathcal{X}\}$. See also Remark 1 for more details.

Proposition 2. *Given two cq-channels* $\mathcal{E}\colon \mathcal{X} \to \mathrm{S}(\mathcal{H}_R)$ *and* $\mathcal{E}'\colon \mathcal{X} \to \mathrm{S}(\mathcal{H}_S)$, *with* $\mathcal{E} = \{\sigma_R^x : x \in \mathcal{X}\}$ *and* $\mathcal{E}' = \{\tau_S^x : x \in \mathcal{X}\}$, *the following are equivalent*:

(i) *For any discrete random variable* U, *any probability distribution* $q(u)$, *and any classical channel* $q(x\,|\,u)$, *the corresponding input-output cq-states*

$$\sigma_{UR} = \sum_{x\in\mathcal{X}} \sum_{u\in\mathcal{U}} q(u)q(x\,|\,u)|u\rangle\langle u|_U \otimes \sigma_R^x \qquad \textit{and} \qquad \tau_{US} = \sum_{x\in\mathcal{X}} \sum_{u\in\mathcal{U}} q(u)q(x\,|\,u)|u\rangle\langle u|_U \otimes \tau_S^x$$

satisfy

$$H_{\min}(U\,|\,R)_\sigma \le H_{\min}(U\,|\,S)_\tau,$$

i.e., $P_{\mathrm{guess}}(U\,|\,R)_\sigma \ge P_{\mathrm{guess}}(U\,|\,S)_\tau$;

(ii) *There exists a Hermitian trace-preserving statistical morphism of* $\mathcal{E}$, *denoted by* $\mathcal{L}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$, *such that*

$$\tau_S^x = \mathcal{L}(\sigma_R^x), \quad x \in \mathcal{X}.$$

4. FIRST EXTENSION RESULT: THE SEMICLASSICAL AND CLASSICAL CASES

One sufficient condition for a statistical morphisms to be extendable to a CPTP map is that the composite map $\mathcal{L} \circ \mathcal{N}$ has commuting output.

Lemma 1. *Let* $\mathcal{L}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ *be a statistical morphism of a channel* $\mathcal{N} \in \mathrm{C}(\mathcal{H}_Q, \mathcal{H}_R)$. *If*

$$\big[(\mathcal{L} \circ \mathcal{N})(\rho)\,(\mathcal{L} \circ \mathcal{N})(\sigma)\big] = 0$$

for all $\rho, \sigma \in \mathrm{S}(\mathcal{H}_Q)$, *then there exists a CPTP map* $\mathcal{T}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ *such that*

$$\mathcal{T} \circ \mathcal{N} = \mathcal{L} \circ \mathcal{N}.$$

Proof. For $\mathcal{L}$ being a statistical morphism of $\mathcal{N}$, we know from Definition 5 that, for any POVM $\{\bar{P}_S^x : x \in \mathcal{X}\}$ on $\mathcal{H}_S$, there exists a POVM $\{P_R^x : x \in \mathcal{X}\}$ on $\mathcal{H}_R$ such that

$$\operatorname{Tr}[(\mathcal{L} \circ \mathcal{N})(\rho_Q)\, \bar{P}_S^x] = \operatorname{Tr}[\mathcal{N}(\rho_Q)\, \mathcal{L}^*(\bar{P}_S^x)] = \operatorname{Tr}[\mathcal{N}(\rho_Q)\, P_R^x],$$

for all $\rho_Q \in \mathrm{S}(\mathcal{H}_Q)$. For $\mathcal{X} = [1, d_S]$, denote by $\{|x\rangle : x \in \mathcal{X}\}$ the orthonormal basis of $\mathcal{H}_S$ that simultaneously diagonalize any output of $\mathcal{L} \circ \mathcal{N}$. (Such a basis exists, since $[(\mathcal{L} \circ \mathcal{N})(\rho), (\mathcal{L} \circ \mathcal{N})(\sigma)] = 0$.) Then choose, in the above equation, $\bar{P}_S^x = |x\rangle\langle x|_S$, and define $\mathcal{T} \colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ to be the linear map given by

$$\mathcal{T}(Z_R) \overset{\text{def}}{=} \sum_{x \in \mathcal{X}} |x\rangle\langle x|_S \operatorname{Tr}[Z_R\, P_R^x],$$

for any $Z_R \in \mathrm{L}(\mathcal{H}_R)$. By the construction, $\mathcal{T}$ is CPTP (indeed, it is a measure-and-prepare quantum channel). Moreover, for all $\rho_Q \in \mathrm{S}(\mathcal{H}_Q)$,

$$\begin{aligned}(\mathcal{T} \circ \mathcal{N})(\rho_Q) &= \sum_{x \in \mathcal{X}} |x\rangle\langle x|_S \operatorname{Tr}[\mathcal{N}(\rho_Q)\, P_R^x] \\ &= \sum_{x \in \mathcal{X}} |x\rangle\langle x|_S \operatorname{Tr}[\mathcal{N}(\rho_Q)\, \mathcal{L}^*(|x\rangle\langle x|_S)] \\ &= \sum_{x \in \mathcal{X}} |x\rangle\langle x|_S \operatorname{Tr}[(\mathcal{L} \circ \mathcal{N})(\rho_Q)\, |x\rangle\langle x|_S] = (\mathcal{L} \circ \mathcal{N})(\rho_Q),\end{aligned}$$

where the last identity comes from the fact that all $\mathcal{N}'(\rho_Q)$ are diagonal on the basis $\{|x\rangle\}$. △

As an immediate consequence of Lemma 1 and Proposition 1, we obtain the following.

Corollary 1. *Let $\mathcal{N} \in \mathrm{C}(\mathcal{H}_Q, \mathcal{H}_R)$ and $\mathcal{N}' \in \mathrm{C}(\mathcal{H}_Q, \mathcal{H}_S)$ be two CPTP maps. Let, moreover, $\mathcal{N}'$ be such that $[\mathcal{N}'(\rho), \mathcal{N}'(\sigma)] = 0$, for all $\rho, \sigma \in \mathrm{S}(\mathcal{H}_Q)$. Then the following are equivalent*:

(i) *For any discrete random variable U, any probability distribution $q(u)$, and any cq-channel $\mathcal{E} = \{\rho_Q^u : u \in \mathcal{U}\}$, the corresponding input-output cq-states*

$$\sigma_{UR} = \sum_{u \in \mathcal{U}} q(u)|u\rangle\langle u|_U \otimes \mathcal{N}_Q(\rho_Q^u) \qquad \textit{and} \qquad \tau_{US} = \sum_{u \in \mathcal{U}} q(u)|u\rangle\langle u|_U \otimes \mathcal{N}'_Q(\rho_Q^u)$$

satisfy

$$H_{\min}(U \,|\, R)_\sigma \le H_{\min}(U \,|\, S)_\tau,$$

i.e., $P_{\text{guess}}(U \,|\, R)_\sigma \ge P_{\text{guess}}(U \,|\, S)_\tau$;

(ii) *$\mathcal{N}$ is degradable into $\mathcal{N}'$, i.e., there exists a CPTP map $\mathcal{T} \colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ such that*

$$\mathcal{N}' = \mathcal{T} \circ \mathcal{N}.$$

The above corollary can be specialized to cq-channels as follows.

Corollary 2. *Consider two cq-channels $\mathcal{E} \colon \mathcal{X} \to \mathrm{S}(\mathcal{H}_R)$ and $\mathcal{E}' \colon \mathcal{X} \to \mathrm{S}(\mathcal{H}_S)$, with $\mathcal{E} = \{\sigma_R^x : x \in \mathcal{X}\}$ and $\mathcal{E}' = \{\tau_S^x : x \in \mathcal{X}\}$. Assume moreover that $[\tau_S^x, \tau_S^{x'}] = 0$, for all $x, x' \in \mathcal{X}$. Then the following are equivalent*:

(i) *For any discrete random variable U, any probability distribution $q(u)$, and any classical channel $q(x \,|\, u)$, the corresponding input-output cq-states*

$$\sigma_{UR} = \sum_{x \in \mathcal{X}} \sum_{u \in \mathcal{U}} q(u) q(x \,|\, u) |u\rangle\langle u|_U \otimes \sigma_R^x \qquad \textit{and} \qquad \tau_{US} = \sum_{x \in \mathcal{X}} \sum_{u \in \mathcal{U}} q(u) q(x \,|\, u) |u\rangle\langle u|_U \otimes \tau_S^x$$

satisfy

$$H_{\min}(U \,|\, R)_\sigma \le H_{\min}(U \,|\, S)_\tau,$$

i.e., $P_{\text{guess}}(U \,|\, R)_\sigma \ge P_{\text{guess}}(U \,|\, S)_\tau$;

(ii) *$\mathcal{E}$ is degradable into $\mathcal{E}'$, i.e., there exists a CPTP map $\mathcal{T} \colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ such that*

$$\tau_S^x = \mathcal{T}(\sigma_R^x), \quad x \in \mathcal{X}.$$

4.1. Classical Case

When both cq-channels have commuting output, we can state the result in purely classical terms as follows.

Corollary 3. *Given two classical noisy channels $p(y\,|\,x)$ and $p'(z\,|\,x)$, the following are equivalent*:

(i) *p is degradable into p'*;

(ii) *For any discrete random variable U, any probability distribution $q(u)$, and any channel $q(x\,|\,u)$, the joint probability distributions $q(u)q(x\,|\,u)p(y\,|\,x)$ and $q(u)q(x\,|\,u)p'(z\,|\,x)$ satisfy*

$$H_{\min}(U\,|\,Y) \le H_{\min}(U\,|\,Z).$$

In other words, by replacing H with $H_{\min}$ in Definition 2, we obtain that the corresponding notion of "less noisy" is equivalent to the notion of "degradable." On the other hand, we recall the fact that there exist less noisy channels that are not degradable [5]. For the reader's convenience, we report below a self-contained proof of Corollary 3, which does not rely on any previous result about statistical morphisms or quantum channels.

Proof of Corollary 3. Obviously, point (i) implies point (ii). Conversely, let us assume (ii). This means that, for any joint probability distribution $q(x, u)$,

$$\max_d \sum_{x,y,u} q(x,u)p(y\,|\,x)d(u\,|\,y) \ge \max_{d'} \sum_{x,z,u} q(x,u)p'(z\,|\,x)d'(u\,|\,z),$$

where $d(u\,|\,y)$ and $d'(u\,|\,z)$ denote the guessing strategies, i.e., discrete noisy channels $d\colon Y \to \widehat{U}$ and $d'\colon Z \to \widehat{U}$, which the receiver can optimize in order to maximize the probability of correct guessing $\Pr\{U = \widehat{U}\}$.

Choose now U with $\mathcal{U} = \mathcal{Z}$, and label its states by z'. Also, fix the strategy $d'(z'\,|\,z) = \delta_{z',z}$. Then, for any $q(x, z')$, there exists $d(z'\,|\,y)$ such that

$$\begin{aligned}\sum_{x,z'} q(x,z')\Bigl(\sum_z p'(z\,|\,x)d'(z\,|\,z') - \sum_y p(y\,|\,x)d(z'\,|\,y)\Bigr)\\ = \sum_{x,z'} q(x,z')\Bigl(p'(z'\,|\,x) - \sum_y p(y\,|\,x)d(z'\,|\,y)\Bigr) \le 0.\end{aligned}$$

Equivalently,

$$\max_q \min_d \left\{\sum_{x,z'} q(x,z')\Bigl(p'(z'\,|\,x) - \sum_y p(y\,|\,x)d(z'\,|\,y)\Bigr)\right\} \le 0.$$

By the minimax theorem (for our case, see [11, Lemma 4.13]) we can exchange the order of the two optimizations, so that

$$\min_d \max_q \left\{\sum_{x,z'} q(x,z')\Bigl(p'(z'\,|\,x) - \sum_y p(y\,|\,x)d(z'\,|\,y)\Bigr)\right\} \le 0.$$

Denoting by $\Delta(x, z')$ the difference $p'(z'\,|\,x) - \sum_y p(y\,|\,x)d(z'\,|\,y)$, we notice that, since $\sum_{x,z'} \Delta(x, z') = 0$, we necessarily have $\max_{x,z'} \Delta(x, z') \ge 0$; otherwise, $\sum_{x,z'} \Delta(x, z') < 0$. Consequently,

$$\min_d \max_q \left\{\sum_{x,z'} q(x,z')\Bigl(p'(z'\,|\,x) - \sum_y p(y\,|\,x)d(z'\,|\,y)\Bigr)\right\} = \min_d \max_{x,z'} \Delta(x,z');$$

i.e., the maximum is achieved by concentrating the probability distribution $q(x, z')$ on one largest entry. Then, for what we said, we know that

$$\min_{d} \max_{x,z'} \Delta(x, z') = 0,$$

implying the existence of a channel $d(z' \,|\, y)$ such that

$$\sum_{y} p(y \,|\, x) d(z' \,|\, y) = p'(z' \,|\, x), \quad \forall x, z';$$

i.e., p is degradable into p', as claimed. △

The main advantage of the above proof, with respect to the one used in the general case, is that it can easily be generalized to the approximate case, namely, when there exists $\varepsilon \geq 0$ such that, for any random variable U and any joint probability distribution $q(x, u)$,

$$P_{\text{guess}}(U \,|\, Y) \geq P_{\text{guess}}(U \,|\, Z) - \varepsilon.$$

This case is studied in the Appendix.

5. SECOND EXTENSION RESULT: THE FULLY QUANTUM CASE

Lemma 2. *For a given CPTP map* $\mathcal{N}\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_R)$ *and a given linear map* $\mathcal{L}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$, *let* $\mathcal{H}_{S'} \cong \mathcal{H}_S$, *and assume that* $(\mathrm{id}_{S'} \otimes \mathcal{L})$ *is a statistical morphism of* $(\mathrm{id}_{S'} \otimes \mathcal{N})$. *Then there exists a CPTP map* $\mathcal{T}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ *such that*

$$\mathcal{T} \circ \mathcal{N} = \mathcal{L} \circ \mathcal{N}.$$

Proof. Since $(\mathrm{id}_{S'} \otimes \mathcal{L})$ is a statistical morphism of $(\mathrm{id}_{S'} \otimes \mathcal{N})$, we know from Definition 5 that, for any POVM $\{\bar{P}^x_{S'S} : x \in \mathcal{X}\}$ on $\mathcal{H}_{S'} \otimes \mathcal{H}_S \cong \mathcal{H}_S^{\otimes 2}$, there exists a POVM $\{P^x_{S'R} : x \in \mathcal{X}\}$ on $\mathcal{H}_{S'} \otimes \mathcal{H}_R$ such that

$$\mathrm{Tr}\big[\{\omega_{S'} \otimes (\mathcal{L} \circ \mathcal{N})(\rho_Q)\}\, \bar{P}^x_{S'S}\big] = \mathrm{Tr}\big[\{\omega_{S'} \otimes (\mathcal{N})(\rho_Q)\}\, P^x_{S'R}\big],$$

for all $x \in \mathcal{X}$, all $\omega_{S'} \in \mathrm{S}(\mathcal{H}_{S'})$, and all $\rho_Q \in \mathrm{S}(\mathcal{H}_Q)$. By linearity, this implies that

$$\mathrm{Tr}_{S'S}\big[\{\Phi^+_{S''S'} \otimes (\mathcal{L} \circ \mathcal{N})(\rho_Q)\}\, \{\mathbb{1}_{S''} \otimes \bar{P}^x_{S'S}\}\big] = \mathrm{Tr}_{S'R}\big[\{\Phi^+_{S''S'} \otimes (\mathcal{N})(\rho_Q)\}\, \{\mathbb{1}_{S''} \otimes P^x_{S'R}\}\big], \qquad (3)$$

for all $x \in \mathcal{X}$ and all $\rho_Q \in \mathrm{S}(\mathcal{H}_Q)$, where $\Phi^+_{S''S'} = d_S \sum\limits_{i,j=1}^{d_S} |i_{S''}\rangle |i_{S'}\rangle \langle j_{S''}| \langle j_{S'}|$ is the maximally entangled state in $\mathrm{S}(\mathcal{H}_{S''} \otimes \mathcal{H}_{S'}) \cong \mathrm{S}(\mathcal{H}_S^{\otimes 2})$.

The protocol of generalized teleportation [21] implies the existence of a POVM $\{B^x_{S''S'} : x \in \mathcal{X}\}$ and unitary operators $\{U^x_{S'' \to S} : x \in \mathcal{X}\}$ such that

$$(\mathcal{L} \circ \mathcal{N})(\rho_Q) = \sum_{x \in \mathcal{X}} U^x_{S'' \to S}\, \mathrm{Tr}_{S'S}\big[\{\Phi^+_{S''S'} \otimes (\mathcal{L} \circ \mathcal{N})(\rho_Q)\}\, \{\mathbb{1}_{S''} \otimes B^x_{S'S}\}\big] (U^x_{S'' \to S})^\dagger,$$

for all $\rho_Q \in \mathrm{S}(\mathcal{H}_Q)$. Then (3) implies the existence of a POVM $\{P^x_{S'R} : x \in \mathcal{X}\}$ on $\mathcal{H}_{S'} \otimes \mathcal{H}_R$ such that

$$(\mathcal{L} \circ \mathcal{N})(\rho_Q) = \sum_{x \in \mathcal{X}} U^x_{S'' \to S}\, \mathrm{Tr}_{S'R}\big[\{\Phi^+_{S''S'} \otimes (\mathcal{N})(\rho_Q)\}\, \{\mathbb{1}_{S''} \otimes P^x_{S'R}\}\big] (U^x_{S'' \to S})^\dagger,$$

for all $\rho_Q \in \mathrm{S}(\mathcal{H}_Q)$. The statement is proved by defining the map $\mathcal{T}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ as

$$\mathcal{T}(Z_R) \stackrel{\text{def}}{=} \sum_{x \in \mathcal{X}} U^x_{S'' \to S}\, \mathrm{Tr}_{S'R}\big[\{\Phi^+_{S''S'} \otimes Z_R\}\, \{\mathbb{1}_{S''} \otimes P^x_{S'R}\}\big] (U^x_{S'' \to S})^\dagger,$$

and noticing that, being a sort of "noisy teleportation," $\mathcal{T}$ is indeed a CPTP map, as claimed. △

In fact, following an argument in [15], it is not difficult to show that the assumption in Lemma 2 can be somewhat weakened as follows: instead of assuming that $(\mathrm{id}_{S'} \otimes \mathcal{L})$ is a statistical morphism of $(\mathrm{id}_{S'} \otimes \mathcal{N})$, one can assume that $(\mathrm{id}_{S'} \otimes \mathcal{L})$ is a statistical morphism of $(\mathcal{D}_{S'} \otimes \mathcal{N})$, where $\mathcal{D}\colon \mathrm{L}(\mathcal{H}_{S'}) \to \mathrm{L}(\mathcal{H}_{S'})$ is some invertible CPTP map, in the sense that $\mathcal{D}(\mathrm{L}(\mathcal{H}_{S'})) = \mathrm{L}(\mathcal{H}_{S'})$. (For example, a channel $\mathcal{D}(\rho) = p\rho + (1-p)\mathbb{1}/d$ is invertible as long as $p > 0$.)

As an immediate consequence of Lemma 2 and Proposition 1, we obtain the following.

Corollary 4. *Let $\mathcal{N}\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_R)$ and $\mathcal{N}'\colon \mathrm{L}(\mathcal{H}_Q) \to \mathrm{L}(\mathcal{H}_S)$ be two CPTP maps. Let $\mathcal{H}_{S'}$ be an auxiliary Hilbert space such that $\mathcal{H}_{S'} \cong \mathcal{H}_S$. The following are equivalent*:

(i) *For any discrete random variable U, any probability distribution $q(u)$, and any cq-channel $\mathcal{E} = \{\rho^u_{S'Q} : u \in \mathcal{U}\}$, the corresponding input-output cq-states*

$$\sigma_{US'R} = \sum_{u\in\mathcal{U}} q(u)|u\rangle\langle u|_U \otimes (\mathrm{id}_{S'} \otimes \mathcal{N}_Q)(\rho^u_{S'Q})$$

and

$$\tau_{US'S} = \sum_{u\in\mathcal{U}} q(u)|u\rangle\langle u|_U \otimes (\mathrm{id}_{S'} \otimes \mathcal{N}'_Q)(\rho^u_{S'Q})$$

satisfy

$$H_{\min}(U\,|\,S'R)_\sigma \le H_{\min}(U\,|\,S'S)_\tau,$$

i.e., $P_{\mathrm{guess}}(U\,|\,S'R)_\sigma \ge P_{\mathrm{guess}}(U\,|\,S'S)_\tau$;

(ii) *$\mathcal{N}$ is degradable into $\mathcal{N}'$, i.e., there exists a CPTP map $\mathcal{T}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ such that*

$$\mathcal{N}' = \mathcal{T} \circ \mathcal{N}.$$

In the case of cq-channels, we have the following.

Corollary 5. *Consider two cq-channels $\mathcal{E}\colon \mathcal{X} \to \mathrm{S}(\mathcal{H}_R)$ and $\mathcal{E}'\colon \mathcal{X} \to \mathrm{S}(\mathcal{H}_S)$, with $\mathcal{E} = \{\sigma^x_R : x \in \mathcal{X}\}$ and $\mathcal{E}' = \{\tau^x_S : x \in \mathcal{X}\}$. Introduce an auxiliary Hilbert space $\mathcal{H}_{S'} \cong \mathcal{H}_S$ and let $\mathcal{E}''\colon \mathcal{Y} \to \mathrm{S}(\mathcal{H}_{S'})$ be a cq-channel, with $\mathcal{E}'' = \{\omega^y_{S'} : y \in \mathcal{Y}\}$, such that $\mathrm{span}\{\omega^y_{S'} : y \in \mathcal{Y}\} = \mathrm{L}(\mathcal{H}_{S'})$. Then the following are equivalent*:

(i) *For any discrete random variable U, any probability distribution $q(u)$, and any classical channel $q(y, x\,|\,u)$, the corresponding input-output cq-states*

$$\sigma_{US'R} = \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}}\sum_{u\in\mathcal{U}} q(u)q(y,x\,|\,u)|u\rangle\langle u|_U \otimes \omega^y_{S'} \otimes \sigma^x_R$$

and

$$\tau_{US'S} = \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}}\sum_{u\in\mathcal{U}} q(u)q(x\,|\,u)|u\rangle\langle u|_U \otimes \omega^y_{S'} \otimes \tau^x_S$$

satisfy

$$H_{\min}(U\,|\,S'R)_\sigma \le H_{\min}(U\,|\,S'S)_\tau,$$

i.e., $P_{\mathrm{guess}}(U\,|\,S'R)_\sigma \ge P_{\mathrm{guess}}(U\,|\,S'S)_\tau$;

(ii) *$\mathcal{E}$ is degradable into $\mathcal{E}'$, i.e., there exists a CPTP map $\mathcal{T}\colon \mathrm{L}(\mathcal{H}_R) \to \mathrm{L}(\mathcal{H}_S)$ such that*

$$\tau^x_S = \mathcal{T}(\sigma^x_R), \quad x \in \mathcal{X}.$$

6. CONCLUSIONS

In this work we have described a connection between the theory of statistical comparison and the comparison of noisy channels, independent of that of [22]. In particular, we have shown how Definitions 1 and 2 become completely equivalent if H is replaced by $H_{\min}$ in Definition 2.

The result proved here can be seen as a *converse* to the data-processing inequality: the monotonic decrease of information (as measured here by $H_{\min}$ or, equivalently, by P_{guess}) is not only necessary but also *sufficient* for the existence of a post-processing map (a trace-preserving statistical morphism, in general, but we saw how additional assumptions can lead to the existence of a CPTP post-processing).

As we have already mentioned in other publications [17, 23–25], we believe that this approach, based on the theory of statistical comparison, can play an important role in understanding the peculiarity of memoryless processes as the information-theoretic counterpart of adiabatic processes in thermodynamics.

APPENDIX

CLASSICAL CASE: APPROXIMATE VERSION

Assuming

$$P_{\text{guess}}(U\,|\,Y) \geq P_{\text{guess}}(U\,|\,Z) - \varepsilon, \quad \varepsilon \geq 0,$$

the proof of Corollary 3 carries through unaltered, until one shows that

$$\min_{d} \max_{x,z'} \Delta(x,z') \leq \varepsilon. \tag{4}$$

To proceed from here, consider now the following quantity:

$$\max_{x} \sum_{z'} |\Delta(x,z')|. \tag{5}$$

The above quantity is the induced ℓ_1-norm distance[3] between the channel $p'(z'\,|\,x)$ and the degraded channel $\sum\limits_{y} p(y\,|\,x)d(z'\,|\,y)$. Since, for all x, $\sum\limits_{z'} \Delta(x,z') = 0$, we have

$$\sum_{z'} |\Delta(x,z')| = 2 \sum_{z'\colon \Delta(x,z')\geq 0} \Delta(x,z'), \quad \forall x \in \mathcal{X},$$

which implies that, for the strategy d achieving the left-hand side of (4),

$$\sum_{z'} |\Delta(x,z')| \leq 2|\mathcal{X}| \max_{x,z'} \Delta(x,z') \leq 2|\mathcal{X}|\varepsilon, \quad \forall x \in \mathcal{X}.$$

In particular,

$$\max_{x} \sum_{z'} \left| p'(z'\,|\,x) - \sum_{y} p(y\,|\,x)d(z'\,|\,y) \right| \leq 2|\mathcal{X}|\varepsilon.$$

We summarize this finding in a separate corollary.

[3] It holds that (see, e.g., [26, Example 5.6.4])

$$\max_{x} \sum_{z'} |\Delta(x,z')| = \max_{x} \sum_{z'} \left| p'(z'\,|\,x) - \sum_{y} p(y\,|\,x)d(z'\,|\,y) \right| \stackrel{\text{def}}{=} |||p' - dp|||_1,$$

where $|||A|||_1 \stackrel{\text{def}}{=} \max\limits_{v\colon \|v\|_1=1} \|Av\|_1$ is the *variational norm.* The quantity in (5) measures how well one can statistically distinguish $p'(z\,|\,x)$ from $\sum\limits_{y} p(y\,|\,x)d(y\,|\,z)$.

Corollary 6. *Given two classical noisy channels*, $p(y\,|\,x)$ *and* $p'(z\,|\,x)$, *and* $\varepsilon \geq 0$, *assume that, for any discrete random variable* U, *any probability distribution* $q(u)$, *and any channel* $q(x\,|\,u)$, *the joint probability distributions* $q(u)q(x\,|\,u)p(y\,|\,x)$ *and* $q(u)q(x\,|\,u)p'(z\,|\,x)$ *satisfy*

$$2^{-H_{\min}(U|Y)} \geq 2^{-H_{\min}(U|Z)} - \varepsilon.$$

Then there exists a degrading channel $d(z\,|\,y)$ *such that*

$$\max_x \sum_z \left| p'(z\,|\,x) - \sum_y p(y\,|\,x)d(z\,|\,y) \right| \leq 2|\mathcal{X}|\varepsilon.$$

REFERENCES

1. Wilde, M.M., *Quantum Information Theory*, Cambridge, UK: Cambridge Univ. Press, 2013.

2. Cohen, J.E., Kemperman, J.H.B., and Zbăganu, G., *Comparisons of Stochastic Matrices, with Applications in Information Theory, Statistics, Economics, and Population Sciences*, Boston: Birkhäuser, 1998.

3. König, R., Renner, R., and Schaffner, C., The Operational Meaning of Min- and Max-Entropy, *IEEE Trans. Inform. Theory*, 2009, vol. 55, no. 9, pp. 4337–4347.

4. Bergmans, P.P., Random Coding Theorem for Broadcast Channels with Degraded Components, *IEEE Trans. Inform. Theory*, 1973, vol. 19, no. 2, pp. 197–207.

5. Körner, J. and Marton, K., Comparison of Two Noisy Channels, *Topics in Information Theory (2nd Colloq., Keszthely, Hungary, 1975)*, Csiszár, I. and Elias, P., Eds., Colloq. Math. Soc. János Bolyai, vol. 16, Amsterdam: North-Holland, 1977, pp. 411–423.

6. El Gamal, A.A., Broadcast Channels with and without Feedback, *Proc. 11th Asilomar Conference on Circuits, Systems and Computers, Pacific Grove, CA, Nov. 7–9, 1977*, Chan, S.-P., Ed., New York: IEEE, 1978, pp. 180–183.

7. Csiszár, I. and Körner, J., Information Theory: Coding Theorems for Discrete Memoryless Systems. Cambridge: Cambridge Univ. Press, 2011, 2nd ed.

8. Watanabe, S., Private and Quantum Capacities of More Capable and Less Noisy Quantum Channels, *Phys. Rev. A*, 2012, vol. 85, no. 1, p. 012326.

9. Blackwell, D., Equivalent Comparisons of Experiments, *Ann. Math. Statist.*, 1953, vol. 24, no. 2, pp. 265–272.

10. Torgersen, E.N., *Comparison of Statistical Experiments*, Cambridge: Cambridge Univ. Press, 1991.

11. Liese, F. and Miescke, K.-J., *Statistical Decision Theory: Estimation, Testing, and Selection*, New York: Springer, 2008.

12. Buscemi, F., Keyl, M., D'Ariano, G.M., Perinotti, P., and Werner, R.F., Clean Positive Operator Valued Measures, *J. Math. Phys.*, 2005, vol. 46, no. 8, p. 82109.

13. Shmaya, E., Comparison of Information Structures and Completely Positive Maps, *J. Phys. A*, 2005, vol. 38, no. 44, pp. 9717–9727.

14. Chefles, A., The Quantum Blackwell Theorem and Minimum Error State Discrimination, `arXiv:0907.0866v4 [quant-ph]`, 2009.

15. Buscemi, F., Comparison of Quantum Statistical Models: Equivalent Conditions for Sufficiency, *Comm. Math. Phys.*, 2012, vol. 310, no. 3, pp. 625–647.

16. Buscemi, F., Datta, N., and Strelchuk, S., Game-Theoretic Characterization of Antidegradable Channels, *J. Math. Phys.*, 2014, vol. 55, no. 9, p. 092202.

17. Buscemi F. and Datta, N., Equivalence between Divisibility and Monotonic Decrease of Information in Classical and Quantum Stochastic Processes, *Phys. Rev. A*, 2016, vol. 93, no. 1, p. 012101.

18. Matsumoto, K., An Example of a Quantum Statistical Model Which Cannot Be Mapped to a Less Informative One by Any Trace Preserving Positive Map, `arXiv:1409.5658 [quant-ph, math.ST]`, 2014.

19. Heinosaari, T., Jivulescu, M.A., Reeb, D., and Wolf, M.M., Extending Quantum Operations, *J. Math. Phys.*, 2012, vol. 53, no. 10, p. 102208.

20. Jenčová, A., Comparison of Quantum Binary Experiments, *Rep. Math. Phys.*, 2012, vol. 70, no. 2, pp. 237–249.

21. Braunstein, S.L., D'Ariano, G.M., Milburn, G.J., and Sacchi, M.F., Universal Teleportation with a Twist, *Phys. Rev. Lett.*, 2000, vol. 84, no. 15, pp. 3486–3489.

22. Raginsky, M., Shannon Meets Blackwell and Le Cam: Channels, Codes, and Statistical Experiments, in *Proc. 2011 IEEE Int. Sympos. on Information Theory (ISIT'2011), St. Petersburg, Russia, July 31 – Aug. 5, 2011*, pp. 1220–1224.

23. Buscemi, F., All Entangled Quantum States are Nonlocal, *Phys. Rev. Lett.*, 2012, vol. 108, no. 20, p. 200401.

24. Buscemi, F., Complete Positivity, Markovianity, and the Quantum Data-Processing Inequality, in the Presence of Initial System-Environment Correlations, *Phys. Rev. Lett.*, 2014, vol. 113, no. 14, p. 140502.

25. Buscemi, F., Fully Quantum Second-Law-like Statements from the Theory of Statistical Comparisons, `arXiv:1505.00535 [quant-ph]`, 2014.

26. Horn, R.A. and Johnson, C.R., *Matrix Analysis*, Cambridge: Cambridge Univ. Press, 2013, 2nd ed.