

9-21-2020

Relative entropy and catalytic relative majorization

Soorya Rethinasamy
Birla Institute of Technology and Science, Pilani

Mark M. Wilde
Louisiana State University

Follow this and additional works at: https://digitalcommons.lsu.edu/physics_astronomy_pubs

Recommended Citation

Rethinasamy, S., & Wilde, M. (2020). Relative entropy and catalytic relative majorization. *Physical Review Research*, 2 (3) <https://doi.org/10.1103/PhysRevResearch.2.033455>

This Article is brought to you for free and open access by the Department of Physics & Astronomy at LSU Digital Commons. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Digital Commons. For more information, please contact ir@lsu.edu.

Relative entropy and catalytic relative majorization

Soorya Rethinasamy^{1,2,*} and Mark M. Wilde^{2,3,†}

¹*Birla Institute of Technology and Science, Pilani, Rajasthan 333031, India*

²*Hearne Institute for Theoretical Physics, Department of Physics and Astronomy, and Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA*

³*Stanford Institute for Theoretical Physics, Stanford University, Stanford, California 94305, USA*



(Received 11 May 2020; accepted 1 September 2020; published 21 September 2020)

Given two pairs of quantum states, a fundamental question in the resource theory of asymmetric distinguishability is whether there exists a quantum channel converting one pair to the other. In this work, we reframe this question in such a way that a catalyst can be used to help perform the transformation, with the only constraint on the catalyst being that its reduced state is returned unchanged, so that it can be used again to assist a future transformation. What we find here, for the special case in which the states in a given pair are commuting, and thus quasiclassical, is that this catalytic transformation can be performed if and only if the relative entropy of one pair of states is larger than that of the other pair. This result endows the relative entropy with a fundamental operational meaning that goes beyond its traditional interpretation in the setting of independent and identical resources. Our finding thus has an immediate application and interpretation in the resource theory of asymmetric distinguishability, and we expect it to find application in other domains.

DOI: [10.1103/PhysRevResearch.2.033455](https://doi.org/10.1103/PhysRevResearch.2.033455)

I. INTRODUCTION

The majorization partial ordering has been studied in light of the following question [1]: What does it mean for one probability distribution to be more disordered than another? While there exist several equivalent characterizations of majorization, the most fundamental is arguably the notion that a distribution p majorizes another distribution p' when p' can be obtained from p by the action of a doubly stochastic channel. This captures the intuition that p' is more disordered than p .

The theory of majorization has several widespread applications, some of which are in quantum resource theories [2]. For example, transformations of pure bipartite states by means of local operations and classical communication can be analyzed in terms of majorization of the Schmidt coefficients of the states [3]. Majorization has been shown to determine whether state transformations are possible not just in the resource theory of entanglement, but also in coherence [4–6] and purity [7] as well.

As it turns out, majorization is a special case of a more general concept introduced in [8–10], which is now called relative majorization [11,12]. This more general notion also has a rich history in the context of statistics, where it goes by the name of “statistical comparison” or “comparison of experiments” [13]. Relative majorization is further generalized by the concept of matrix majorization, as presented in [14],

while the d -majorization of [15] is a special case of relative majorization.

To recall the notion of relative majorization, a pair (p, q) of probability distributions p and q relatively majorizes another pair (p', q') of probability distributions p' and q' if there exists a classical channel (conditional probability distribution) that converts p to p' and q to q' . That is, (p, q) relatively majorizes (p', q') , denoted by

$$(p, q) \succ (p', q'), \quad (1)$$

if there exists a classical channel or stochastic matrix N such that $p' = Np$ and $q' = Nq$. Deciding relative majorization can be accomplished by means of a linear program [14] that is efficient in the alphabet sizes of the distributions. In each pair (p, q) and (p', q') , if the second distributions q and q' are equal, then relative majorization collapses to d -majorization. If they are furthermore set to the uniform distribution, then d -majorization collapses to majorization, demonstrating that relative majorization is indeed a generalization of majorization.

Another generalization of majorization comes in the form of catalysis [16]. A catalyst is an ancillary distribution whose presence enables certain transformations that would otherwise be impossible, with the only constraint being that its reduced state is returned unchanged. Catalysis has found application in various resource theories, including entanglement [16,17], thermodynamics [18], purity [19], and coherence [20,21].

In this paper, we reformulate the transformation task in relative majorization by allowing for a catalyst that can aid the transformation. In a general form for this task, a catalyst consists of a pair (r, s) of probability distributions r and s . The pair can be used in conjunction with the original input pair (p, q) to generate the output pair (p', q') by means of a classical channel Λ acting on $(p \otimes r, q \otimes s)$. To make the

*sooryarethin@gmail.com

†mwilde@lsu.edu

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

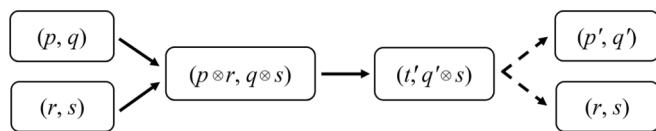


FIG. 1. (p, q) is a pair of probability distributions to be converted to another pair (p', q') by means of a classical channel and the use of the catalyst pair (r, s) . The distribution t' after transformation has first marginal p' and second marginal r . The dashed arrows on the right indicate marginalization of the joint probability distributions t' and $q' \otimes s$.

catalytic task nontrivial, we demand that the first and second marginals of $\Lambda(p \otimes r)$ are p' and r , respectively. We also demand for $\Lambda(q \otimes s)$ to be exactly equal to $q' \otimes s$. As a consequence of this demand, the catalyst is returned unchanged and can be used in future catalytic tasks for distributions that are independent of p and p' [22]. We call this task *relative majorization assisted by a catalyst with correlations*. This task is succinctly summarized in Fig. 1.

We now summarize one result of our paper. We note that we focus on a special case of the task mentioned above, in which the second distribution s of the catalyst pair is fixed to be the uniform distribution, denoted by η . Let p, q, p' , and q' be probability distributions over the same alphabet. Suppose that q and q' have rational entries and full support. Define the *relative spectrum* of the pair (p, q) of probability distributions p and q to be the set formed from the distinct entries of the following set:

$$\{p(x)/q(x)\}_x. \quad (2)$$

Suppose further that the relative spectra of the pairs (p, q) and (p', q') are different. Then, under these assumptions on p, q, p' , and q' , it is possible to transform the pair (p, q) to the pair (p', q') in the above sense if and only if

$$D(p\|q) > D(p'\|q') \quad \text{and} \quad D_0(p\|q) \geq D_0(p'\|q'), \quad (3)$$

where the relative entropy $D(p\|q)$ is defined as

$$D(p\|q) := \sum_x p(x) \ln \left(\frac{p(x)}{q(x)} \right), \quad (4)$$

and the min-relative entropy $D_0(p\|q)$ is defined as

$$D_0(p\|q) := -\ln \left[\sum_{x:p(x) \neq 0} q(x) \right]. \quad (5)$$

Observe that $D_0(p\|q) \geq 0$ for all distributions p and q , and $D_0(p\|q) = 0$ if p has full support, so that the min-relative entropy constraint in (3) can be interpreted as a technical condition.

Another contribution of our paper is that it is possible to perform an inexact yet arbitrarily accurate transformation with fewer assumptions on p, q, p' , and q' , but again under the assumption that s is the uniform distribution.

We establish these results by building on the prior work of [18,23]. As done in those prior works, we employ an embedding channel as a technical tool [18] (see also Ref. [10], p. 227), and we also allow for the catalyst to become correlated with the target distribution [23].

We note here that our results apply to the more general “quasi-classical” case discussed in the abstract, in which the first pair consists of commuting quantum states ρ and σ , and the second pair consists of commuting quantum states ρ' and σ' . This follows as a direct consequence of the fact that commuting quantum states are diagonal in the same basis and thus effectively classical, along with the fact that there is a local unitary channel that takes the common basis of the first pair to the common basis of the second pair.

II. MAIN RESULT

We begin by stating the main results of our paper and an associated corollary. We present all proofs in Appendices C, D, and E, while Appendices A and B review some basics needed to prove the main results.

Theorem 1. One-shot characterization of exact pair transformations. Let p, q, p', q' be probability distributions on the same alphabet, such that the relative spectra of the pairs (p, q) and (p', q') are different, and q and q' have full support and only rational entries. Then the following conditions are equivalent:

$$(i) D(p\|q) > D(p'\|q') \quad \text{and} \quad D_0(p\|q) \geq D_0(p'\|q').$$

(ii) For all $\gamma > 0$, there exists a probability distribution r , a joint distribution t' , and a classical channel Λ such that (a) $\Lambda(p \otimes r) = t'$ with marginals p' and r ; (b) $\Lambda(q \otimes \eta) = q' \otimes \eta$; and (c) $D(t'\|p' \otimes r) < \gamma$; where η is the uniform distribution on the support of r .

We note that statement (ii) above can alternatively be written in the notation of (1) as follows: For all $\gamma > 0$, there exists a probability distribution r and a joint distribution t' such that

$$(p \otimes r, q \otimes \eta) \succ (t', q' \otimes \eta), \quad (6)$$

$$D(t'\|p' \otimes r) < \gamma, \quad (7)$$

where t' has first marginal p' and second marginal r , and η is the uniform distribution on the support of r .

Theorem 1 indicates that the relative entropy is the main relevant information quantity that characterizes pair transformations when catalysts are available, other than the min-relative entropy as an additional technical condition in the case in which p' does not have full support. Furthermore, since $\gamma > 0$ is arbitrary, the relative entropy $D(t'\|p' \otimes r)$ can be made as small as desired, and thus the output of the classical channel Λ can be arbitrarily close to the product of p' and r .

The constraint is that the relative spectra are different, that q and q' are rational, and the min-relative entropy inequality can be relaxed by allowing for a slight error in the formation of the target state, as follows:

Theorem 2. One-shot characterization of approximate pair transformations. Given probability distributions p, q, p', q' on the same alphabet, such that q and q' have full support, the following conditions are equivalent:

$$(i) D(p\|q) \geq D(p'\|q').$$

(ii) For all $\varepsilon \in (0, 1)$ and $\gamma > 0$, there exist probability distributions r and p'_ε , a joint distribution t'_ε , and a classical channel Λ such that (a) $\Lambda(p \otimes r) = t'_\varepsilon$ with marginals p'_ε and r ; (b) $\Lambda(q \otimes \eta) = q' \otimes \eta$; (c) $\frac{1}{2} \|p' - p'_\varepsilon\|_1 \leq \varepsilon$; and (d) $D(t'_\varepsilon\|p'_\varepsilon \otimes r) < \gamma$; where $\|w - v\|_1 := \sum_i |w_i - v_i|$ and η is the uniform distribution on the support of r .

We again note that statement (ii) above can alternatively be written in the notation of (1) as follows: For all $\varepsilon \in (0, 1)$ and $\gamma > 0$, there exist probability distributions r and p'_ε and a joint distribution t'_ε such that

$$(p \otimes r, q \otimes \eta) \succ (t'_\varepsilon, q' \otimes \eta), \quad (8)$$

$$\frac{1}{2} \|p' - p'_\varepsilon\|_1 \leq \varepsilon, \quad (9)$$

$$D(t'_\varepsilon \| p'_\varepsilon \otimes r) < \gamma, \quad (10)$$

where t'_ε has first marginal p'_ε and second marginal r , and η is the uniform distribution on the support of r .

The relative entropy $D(t'_\varepsilon \| p'_\varepsilon \otimes r)$ can be made as small as desired, and thus the output of the classical channel Λ can be arbitrarily close to the product of p'_ε and r . Furthermore, the state p'_ε can be arbitrarily close to the target distribution p' in normalized trace distance. Note that allowing for a slight error in the transformation of p while demanding no error in the transformation of q is consistent with the frameworks of [24–26], with the framework of [25] being known as the “resource theory of asymmetric distinguishability” due to this asymmetry in the transformation.

We now consider a special case of Theorem 2, where we fix both q and q' to be the uniform distribution. A closely related yet different claim is given as follows (Ref. [27], Lemma 3.10):

Corollary 1. Given probability distributions p and p' on the same alphabet such that $p \neq Pp'$ for all permutation matrices P , the following conditions are equivalent:

(i) $H(p) \leq H(p')$.

(ii) For all $\varepsilon \in (0, 1)$ and $\gamma > 0$, there exist probability distributions r and p'_ε , a joint distribution t'_ε , and a unital classical channel Λ such that (a) $\Lambda(p \otimes r) = t'_\varepsilon$ with marginals p'_ε and r , (b) $\frac{1}{2} \|p' - p'_\varepsilon\|_1 \leq \varepsilon$, and (c) $D(t'_\varepsilon \| p'_\varepsilon \otimes r) < \gamma$.

A unital classical channel is one that preserves the uniform distribution. The corollary states that, given a catalyst, the feasibility condition for the transformation is based exclusively on the Shannon entropy. Furthermore, the relative entropy $D(t'_\varepsilon \| p'_\varepsilon \otimes r)$ can be made as small as desired, and thus the output of the classical channel Λ can be arbitrarily close to the product of p'_ε and r . Lastly, the state p'_ε can be arbitrarily close to the target distribution p' in normalized trace distance.

We note here that Corollary 1 applies more generally to quantum states ρ and ρ' of the same dimension and with different spectra, where the first statement involves the von Neumann entropy and the second statement has Λ become a quantum channel that incorporates a local unitary performing a change of basis from the eigenbasis of ρ to the eigenbasis of ρ' .

III. DISCUSSIONS AND IMPLICATIONS

In this section, we discuss various implications of our results. A catalyst is a special ancillary probability distribution whose presence enables certain transformations that would otherwise be impossible. As a consequence, broadening the class of allowed transformations in turn broadens the set of input and output distributions satisfying the mathematical condition governing the feasibility of a transformation [16]. Additionally, allowing for the catalyst to be correlated

with the target distribution, subject to the condition that its reduced state is returned unchanged, previously impossible transformations become possible. Since the catalyst is returned unchanged, the paradigm of catalytic majorization with correlations is physically relevant and justified. In this latter case, further broadening the class of allowed transformations even simplifies the mathematical condition governing the feasibility of a transformation to a single entropic condition.

In the resource theories of pure-state entanglement, purity, and pure-state coherence, state transformations in the absence of a catalyst are governed by the majorization partial order, whereas in the resource theory of quasiclassical thermodynamics, transformations are governed by thermomajorization [28]. The catalytic majorization condition is equivalent to an infinite set of conditions on the Rényi entropy [29,30]. The catalytic majorization condition is not as strict as the majorization condition, as there exist distributions p and q for which p does not majorize q but p catalytically majorizes q [16], establishing that catalysis provides an advantage.

As mentioned above, catalytic majorization can be further relaxed. In catalytic majorization, the target state, after transformation, is required to be in a product state with the catalyst, i.e., the catalyst and the target state are independent after the transformation. Relaxing this additional assumption and allowing the final state to be a nonproduct state, with the condition that the local states remain unchanged, further broadens the allowed transformations and simplifies the mathematical condition governing the feasibility of a transformation. This is a generalization of catalytic majorization, called catalytic majorization with correlations, as a product state is a special case of a general distribution.

Our result in Theorem 2 is a necessary and sufficient condition for the last case, where we allow catalysis and correlations. The result is a single entropic condition involving relative entropy, and it neatly ties together catalytically correlated transformations in the resource theories of asymmetric distinguishability, quasiclassical thermodynamics, and purity into a single paradigm, due to the generality of our setting. Our result also introduces a different operational meaning for the relative entropy by establishing its relevance in the single-shot regime.

IV. RELATED WORK

There are some regimes of operating that are complementary to the ones given in our paper. This section provides a brief overview of these regimes and recalls some known results.

Classical relative majorization in (1) is known to be equivalent to the hypothesis testing region of (p', q') being contained in the hypothesis testing region of (p, q) [13] (see also, e.g., [14], where the region is called a “zonotope,” or [11] with motivations coming from quantum resource theories). Quantum relative majorization was defined recently in [12] as an intended generalization of the classical relative majorization concept in (1), in which we write $(\rho, \sigma) \succ (\rho', \sigma')$, where ρ, σ, ρ' , and σ' are quantum states, to indicate that the hypothesis testing region for (ρ', σ') lies inside that of (ρ, σ) . Other equivalent conditions are given in [12]. In the case of qubit systems, the condition $(\rho, \sigma) \succ (\rho', \sigma')$ is equivalent to there

TABLE I. Results in various regimes and resource theories: the table entries feature conditions for determining whether a transformation is possible in a given regime and resource theory. In all cases, the initial resource is unprimed and labeled by p or (p, q) , and the target resource is primed and labeled by p' or (p', q') . The bottom right cell with a bold entry indicates our contribution. The one-shot regime with catalysis depends on H_α or D_α for all $\alpha \in [-\infty, \infty]$. In the one-shot cases, by broadening the class of transformations allowed, more distributions can satisfy the mathematical conditions governing the feasibility of a transformation. When catalytic transformations with correlations are allowed, the mathematical condition governing the feasibility of a transformation becomes a simple entropic condition. In the various resource theories listed in the table, p refers to different objects: for entanglement theory, it denotes the Schmidt coefficients; for coherence, the incoherent coefficients; and for thermodynamics, the block-diagonal energy coefficients (see, e.g., [33]). We have placed question marks in the table entries that have not been explored (to the best of our knowledge).

Resource theory/regime	Asymptotic	One-shot	One-shot + catalysis	One-shot + catalysis + corr.
Entanglement	$\frac{H(p)}{H(p')}$ [31]	$p < p'$ [3]	$H_\alpha(p) \geq H_\alpha(p')$ [29,30]	?
Coherence	$\frac{H(p)}{H(p')}$ [6]	$p < p'$ [4]	$H_\alpha(p) \geq H_\alpha(p')$ [20]	?
Thermodynamics	$\frac{D(p \gamma)}{D(p' \gamma)}$ [32]	$p \succ_{\text{thermo}} p'$ [28]	$F_\alpha(p) \geq F_\alpha(p')$ [18]	$F(p) \geq F(p')$ [23]
Distinguishability	$\frac{D(p q)}{D(p' q')}$ [25,26]	$(p, q) \succ (p', q')$ [13,14]	$D_\alpha(p q) \geq D_\alpha(p' q')$ [18]	$D(p q) \geq D(p' q')$

existing a quantum channel \mathcal{N} such that [34]

$$\mathcal{N}(\rho) = \rho', \tag{11}$$

$$\mathcal{N}(\sigma) = \sigma', \tag{12}$$

which is consistent with (1). As stated above, for the fully classical case, which has apparently motivated much of the related work in quantum information, the containment of testing regions is equivalent to (1) [13]. However, the equivalence no longer holds as soon as the dimension of the input states is 3, with the dimension of the output states still being 2 [35]. Regardless, given states ρ, σ, ρ' , and σ' , the existence of a quantum channel \mathcal{N} satisfying (11) and (12) can be decided by means of a semidefinite program [36] (see also [25]). This latter paper [36] also developed a more general notion of majorization, called quantum majorization, which is based on (11) and (12), but it generalizes it further.

Majorization and relative majorization are most prominent in the single-shot regime, in which there is no assumed structure on the states involved. There exists another regime that deals with situations in which many independent and identically distributed (i.i.d.) systems are available, often referred to as the asymptotic regime. In the asymptotic regime, perhaps the more pertinent question is that of the rate of conversion from initial to target distributions and not necessarily the feasibility of the transformation. For the resource theories of pure-state entanglement, pure-state coherence, and quasiclassical thermodynamics, the rate of conversion is given by ratios of the entropy (pure-state entanglement and pure-state coherence) and relative entropy to the Gibbs distribution (quasiclassical thermodynamics). The asymptotic version of Blackwell’s “comparison of experiments” paradigm was studied in [24–26]. In [25], the problem was approached by introducing the resource theory of asymmetric distinguishability, which allowed for analyzing this problem in a resource-theoretic way. The result of [25,26] is that, in the asymptotic regime, the rate of conversion between two pairs of probability distributions (p, q) and (p', q') is given by the ratio of the relative entropy of the pairs, thus enhancing the fundamental operational meaning of the relative entropy. This asymptotic result generalizes that from [32], which focused on the resource theory of thermodynamics in which the second

states in the initial and final pairs are set to the thermal states.

Some prior results in various regimes and resource theories are presented in Table I. As indicated in our paper and the table, our results are relevant for the single-shot regime in the context of relative catalytic majorization with correlations, and they are thus complementary to this prior work.

V. CONCLUSION

The majorization partial order was developed to capture the notion of disorder in probability distributions, and it governs transformations in various resource theories. The addition of a catalyst allows for transformations that are not possible in the framework of majorization without catalysis. Furthermore, allowing correlations between the target and the catalyst further broadens the class of allowed transformations and simplifies the feasibility condition for a transformation to a simple entropic condition.

The main result of our paper is that a pair (p, q) of probability distributions can be converted approximately to another pair (p', q') of probability distributions assisted by a catalyst with correlations if and only if $D(p||q) \geq D(p' || q')$. As mentioned earlier, we focus on the special case in which the second distribution s of the catalyst pair is set to the uniform distribution. Our result thus extends the operational meaning of the relative entropy beyond the traditional i.i.d. regime to the single-shot regime. This finding thus complements and enhances the previous finding from [19] in the context of entropy and the resource theory of purity.

There are several open questions yet to be tackled. The most pertinent is whether our result applies to pairs of general quantum states (ρ, σ) and (τ, ω) and quantum catalysts. This appears to be a very challenging question that will likely require new techniques. It would also be interesting to generalize the main results to continuous probability density functions, and it seems that the methods of [10] should be helpful in this regard. Another avenue to be explored is whether our result applies to larger sets of distributions, and not just pairs, i.e., the conversion of the set $\{p_1, p_2, \dots, p_k\}$ of k probability distributions to another set $\{q_1, q_2, \dots, q_k\}$ of k probability distributions for $k > 2$. Lastly, another open question is whether the second element η of the catalyst in

Theorems 1 and 2 can feature an arbitrary probability distribution instead of the uniform distribution.

ACKNOWLEDGMENTS

We thank Francesco Buscemi, Vishal Katariya, and Aby Philip for discussions. We are grateful to the anonymous referees for comments that helped to improve our paper. We acknowledge support from NSF Grant No. 1907615. S.R. thanks the Department of Physics and Astronomy at Louisiana State University for hospitality during his research visit in Fall 2019. S.R. also thanks the Department of Physics, Birla Institute of Technology and Science for this opportunity. M.M.W. acknowledges support from Stanford QFARM and AFOSR (FA9550-19-1-0369).

APPENDIX A: NOTATIONS AND DEFINITIONS

We review various distinguishability measures and their properties in this Appendix. Furthermore, we recall the definition of majorization and the embedding channel, which are used in the forthcoming theorems and lemmas.

1. Rényi divergence

For probability distributions $p = \{p_i\}_{i=1}^k$ and $q = \{q_i\}_{i=1}^k$, the Rényi divergence is defined for all $\alpha \in \mathbb{R} \setminus \{0, 1, \infty, -\infty\}$ as follows [37]:

$$D_\alpha(p\|q) := \frac{\text{sgn}(\alpha)}{\alpha - 1} \ln \sum_{i=1}^k p_i^\alpha q_i^{1-\alpha}, \tag{A1}$$

where

$$\text{sgn}(\alpha) := \begin{cases} 1, & \alpha \geq 0, \\ -1, & \alpha < 0. \end{cases} \tag{A2}$$

The Rényi divergence can alternatively be written as follows:

$$D_\alpha(p\|q) = \frac{\text{sgn}(\alpha)}{\alpha - 1} \ln \sum_{i=1}^k p_i \left(\frac{p_i}{q_i}\right)^{\alpha-1} \tag{A3}$$

$$= \frac{\text{sgn}(\alpha)}{\alpha - 1} \ln \sum_{i=1}^k q_i \left(\frac{p_i}{q_i}\right)^\alpha, \tag{A4}$$

which allows one to think of the expressions inside the logarithm as averages of tilted likelihood ratios. Note that when $\alpha < 0$, the following identity holds:

$$D_\alpha(p\|q) = \frac{|\alpha|}{|\alpha| + 1} D_{|\alpha|+1}(q\|p), \tag{A5}$$

so that the properties of $D_\alpha(p\|q)$ for negative values of α can be deduced from the properties of $D_\alpha(p\|q)$ for values of $\alpha > 1$.

For $\alpha \in \{0, 1, \infty, -\infty\}$, we evaluate $D_\alpha(p\|q)$ as a limit to these values, and we arrive at the following expressions:

$$D_0(p\|q) = -\ln \sum_{i=1: p_i \neq 0}^k q_i, \tag{A6}$$

$$D_1(p\|q) = \sum_{i=1}^k p_i \ln \left(\frac{p_i}{q_i}\right), \tag{A7}$$

$$D_\infty(p\|q) = \ln \max_i \frac{p_i}{q_i}, \tag{A8}$$

$$D_{-\infty}(p\|q) = D_\infty(q\|p). \tag{A9}$$

Note that $D_\infty(p\|q)$ can also be expressed as

$$D_\infty(p\|q) = \ln \min \left\{ \lambda | \lambda \geq \frac{p_i}{q_i} \forall i \right\}. \tag{A10}$$

An important property of the Rényi divergence is that it obeys the data-processing inequality [38], i.e., for all classical channels Λ and $\alpha \in [-\infty, \infty]$ the following inequality holds:

$$D_\alpha(p\|q) \geq D_\alpha(\Lambda(p)\|\Lambda(q)). \tag{A11}$$

Another important property of the Rényi divergence is that, for $\alpha \in [-\infty, \infty]$, it is non-negative,

$$D_\alpha(p\|q) \geq 0 \tag{A12}$$

for all probability distributions p and q , and for all $\alpha \in [-\infty, \infty] \setminus \{0\}$ it is equal to 0,

$$D_\alpha(p\|q) = 0 \tag{A13}$$

if and only if $p = q$. For $\alpha = 0$, it is equal to 0 if $p = q$, but the converse is not generally true.

The Rényi divergence is additive for product probability distributions [38], i.e., for probability distributions p_1, p_2, q_1 , and q_2 , the following equality holds:

$$D_\alpha(p_1 \otimes p_2 \| q_1 \otimes q_2) = D_\alpha(p_1 \| q_1) + D_\alpha(p_2 \| q_2). \tag{A14}$$

2. Rényi entropy

For a probability distribution $p = \{p_i\}_{i=1}^k$, the Rényi entropy is defined for all $\alpha \in \mathbb{R} \setminus \{0, 1, \infty, -\infty\}$ as follows [37]:

$$H_\alpha(p) := \frac{\text{sgn}(\alpha)}{1 - \alpha} \ln \sum_{i=1}^k p_i^\alpha. \tag{A15}$$

Similarly to the Rényi divergence, we also define the Rényi entropy for $\alpha \in \{0, 1, \infty, -\infty\}$ as a limit to these values, and the result is the following expressions:

$$H_0(p) = \ln \text{rank}(p), \tag{A16}$$

$$H_1(p) = -\sum_{i=1}^k p_i \ln p_i, \tag{A17}$$

$$H_\infty(p) = -\ln p_{\max}, \tag{A18}$$

$$H_{-\infty}(p) = \ln p_{\min}. \tag{A19}$$

The following equality relates the Rényi divergence to the Rényi entropy for all $\alpha \in [-\infty, \infty]$:

$$D_\alpha(p\|\eta_k) = \ln k - H_\alpha(p), \tag{A20}$$

where η_k is the uniform distribution on the alphabet of p .

3. Quantum Rényi divergence

The Rényi divergence from the classical regime can be extended to the quantum regime, but the extension is not unique.

Classical probability distributions can be modeled as quantum density operators that are diagonal in a fixed orthonormal basis. Thus, modeling two probability distributions is equivalent to working with two density operators that commute with each other, i.e., that are diagonal in the same basis. However, quantum states do not generally commute, which leads to the nonuniqueness of the quantum extension. Note that any quantum extension should collapse to the classical case if the states commute.

Here, we employ the following quantum Rényi divergence [39]:

$$D_\alpha(\rho\|\sigma) := \frac{\text{sgn}(\alpha)}{\alpha - 1} \ln \text{Tr}(\rho^\alpha \sigma^{1-\alpha}). \tag{A21}$$

The Rényi divergence D_α converges to the following expressions in the limit as α tends to either 0 or 1:

$$D_0(\rho\|\sigma) = -\ln \text{Tr}(\Pi_\rho \sigma), \tag{A22}$$

$$D_1(\rho\|\sigma) = \text{Tr}(\rho(\ln \rho - \ln \sigma)), \tag{A23}$$

where Π_ρ is the projector onto the support of ρ . The quantum Rényi divergence obeys the data-processing inequality for all quantum channels Λ and $\alpha \in [0, 2]$ [39]:

$$D_\alpha(\rho\|\sigma) \geq D_\alpha(\Lambda(\rho)\|\Lambda(\sigma)). \tag{A24}$$

Another property of the quantum Rényi divergence is that for states ρ and σ [39], it is non-negative,

$$D_\alpha(\rho\|\sigma) \geq 0, \tag{A25}$$

and for all $\alpha \in [-\infty, \infty] \setminus \{0\}$, it is equal to 0,

$$D_\alpha(\rho\|\sigma) = 0, \tag{A26}$$

if and only if $\rho = \sigma$. If $\alpha = 0$, it is equal to 0 if $\rho = \sigma$ but the converse is not generally true.

4. Majorization and its extensions

The theory of majorization was developed to capture the notion of disorder. The crux of the theory is to answer the following question: When is one probability distribution more disordered than another? The results can be succinctly put forth as follows [1].

For two probability distributions $p, q \in \mathbb{R}^k$, we say that “ p majorizes q ” or $p \succ q$ if for all $n \in \{1, \dots, k - 1\}$,

$$\sum_{i=1}^n p_i^\downarrow \geq \sum_{i=1}^n q_i^\downarrow, \tag{A27}$$

$$\sum_{i=1}^k p_i^\downarrow = \sum_{i=1}^k q_i^\downarrow, \tag{A28}$$

where p^\downarrow is the reordering of p in descending order. If $p \succ q$, it follows that q is more disordered than p in a precise sense stated in Lemma 1 in Appendix B. That is, there are several formulations of majorization that are known to be equivalent.

An extension of the theory of majorization comes in the form of catalysis. Given two distributions p and q such that $p \not\succeq q$, does there exist a distribution r such that $p \otimes r \succ q \otimes r$? If such a distribution exists, it can be used to transform

p into q and be left unchanged. Since r is unaffected by the process, we call it the catalyst of the transformation. We say that q is catalytically majorized by p , written $p \succ_T q$, if there exists a catalyst r such that $p \otimes r \succ q \otimes r$. The theory of catalytic majorization is not as well understood as majorization. However, several important results have been established [29,30].

5. Embedding channel

A mathematical tool frequently used in the resource theory of thermodynamics is an embedding channel, which is a classical channel that maps a thermal distribution to a uniform distribution [18]. We use this tool in a more general sense needed in our context here in order to map a probability distribution described by a set of rational numbers to a uniform distribution. Note that the embedding channel was originally introduced in Ref. [10] (p. 227) for continuous probability density functions.

Consider the simplex $\{p_i\}_{i=1}^k$ of probability distributions and a set $d := \{d_i\}_{i=1}^k$ of natural numbers with

$$N := \sum_{i=1}^k d_i. \tag{A29}$$

The embedding channel $\Gamma_d : \mathbb{R}^k \rightarrow \mathbb{R}^N$, corresponding to the set d , is defined as

$$\Gamma_d(p) := \bigoplus_{i=1}^k p_i \eta_i \tag{A30}$$

$$= \left(\underbrace{\frac{p_1}{d_1}, \dots, \frac{p_1}{d_1}}_{d_1}, \dots, \underbrace{\frac{p_k}{d_k}, \dots, \frac{p_k}{d_k}}_{d_k} \right), \tag{A31}$$

where $\eta_i = (\frac{1}{d_i}, \dots, \frac{1}{d_i}) \in \mathbb{R}^{d_i}$ is the uniform distribution.

There exists a left inverse for the embedding channel Γ_d , which is denoted by $\Gamma_d^* : \mathbb{R}^N \rightarrow \mathbb{R}^k$ and is defined as follows:

$$\Gamma_d^*(x) := \left(\sum_{i=1}^{d_1} x_i, \sum_{i=d_1+1}^{d_1+d_2} x_i, \dots, \sum_{i=d_1+\dots+d_{n-1}+1}^N x_i \right). \tag{A32}$$

Note that Γ_d^* is itself a classical channel, and furthermore, it reverses the action of Γ_d in the following sense:

$$\Gamma_d^* \circ \Gamma_d = \text{id}, \tag{A33}$$

where id is the identity channel.

APPENDIX B: TECHNICAL LEMMAS

Several of the technical lemmas presented in this Appendix have been established in prior work. Here we list them, with modifications and extensions, for convenience and completeness.

Lemma 1 (Majorization [1,40]). Let $x, y \in \mathbb{R}^k$ be probability distributions. Then the following are equivalent:

- (i) $x \succ y$.

(ii) For $n \in \{1, \dots, k - 1\}$,

$$\sum_{i=1}^n x_i^\downarrow \geq \sum_{i=1}^n y_i^\downarrow, \tag{B1}$$

$$\sum_{i=1}^k x_i^\downarrow = \sum_{i=1}^k y_i^\downarrow. \tag{B2}$$

(iii) $y = Dx$ for some $k \times k$ doubly stochastic matrix D .

(iv) For all $t \in \mathbb{R}$, $\sum_{i=1}^k |x_i - t| \geq \sum_{i=1}^k |y_i - t|$.

Lemma 2 (Embedding of distribution with rational entries [18]). Let γ_d be a probability distribution defined as follows:

$$\gamma_d := \left(\frac{d_1}{N}, \dots, \frac{d_k}{N} \right), \tag{B3}$$

where $d := \{d_i\}_{i=1}^k$ is a set of natural numbers such that

$$N := \sum_{i=1}^k d_i. \tag{B4}$$

Then the action of the embedding channel Γ_d on γ_d is as follows:

$$\Gamma_d(\gamma_d) = \eta_N, \tag{B5}$$

where η_N is the uniform distribution of size N .

Proof. From the definition of an embedding channel Γ_d corresponding to the same set d (see Appendix A 5), we find that

$$\begin{aligned} \Gamma_d(\gamma_d) &= \left(\underbrace{\frac{d_1/N}{d_1}, \dots, \frac{d_1/N}{d_1}}_{d_1}, \dots, \underbrace{\frac{d_k/N}{d_k}, \dots, \frac{d_k/N}{d_k}}_{d_k} \right) \\ &= \left(\underbrace{1/N, \dots, 1/N}_{d_1}, \dots, \underbrace{1/N, \dots, 1/N}_{d_k} \right) \\ &= \left(\underbrace{1/N, \dots, 1/N}_N \right) \\ &= \eta_N. \end{aligned} \tag{B6}$$

This concludes the proof. \blacksquare

Lemma 3 (Preservation of Rényi divergence [18]). Let $p := \{p_i\}_{i=1}^k$ be a probability distribution and let $d := \{d_i\}_{i=1}^k$ be a set of natural numbers with $N := \sum_{i=1}^k d_i$. Let \tilde{p} denote the distribution $\tilde{p} := \Gamma_d(p)$. Then for all $\alpha \in [-\infty, \infty]$, the following equality holds:

$$D_\alpha(p \parallel \gamma_d) = D_\alpha(\tilde{p} \parallel \eta_N), \tag{B7}$$

where γ_d is as defined in Lemma 2 and Γ_d is as defined in Appendix A 5.

Proof. This is a consequence of the data-processing inequality in (A11) holding for all $\alpha \in [-\infty, \infty]$, as well as the identity in (A33). Indeed, for $\alpha \in [-\infty, \infty]$, the following inequality holds as a consequence of data processing:

$$D_\alpha(p \parallel \gamma_d) \geq D_\alpha(\Gamma_d(p) \parallel \Gamma_d(\gamma_d)) = D_\alpha(\tilde{p} \parallel \eta_N). \tag{B8}$$

Additionally, the following inequality is a consequence of data processing and the identity in (A33):

$$D_\alpha(\tilde{p} \parallel \eta_N) = D_\alpha(\Gamma_d(p) \parallel \Gamma_d(\gamma_d)) \tag{B9}$$

$$\geq D_\alpha((\Gamma_d^* \circ \Gamma_d)(p) \parallel (\Gamma_d^* \circ \Gamma_d)(\gamma_d)) \tag{B10}$$

$$= D_\alpha(p \parallel \gamma_d). \tag{B11}$$

This concludes the proof. \blacksquare

Lemma 4 (Mixture reduces Rényi divergence [18]). Let p and q be probability distributions such that $p \neq q$ and q is full rank. Then $\forall \alpha \in (-\infty, \infty) \setminus \{0\}$ and $0 < \delta < 1$,

$$D_\alpha((1 - \delta)p + \delta q \parallel q) < D_\alpha(p \parallel q). \tag{B12}$$

Proof. We split the proof into separate parts for different α values. For $\alpha \in (0, 1]$, the following inequality is a consequence of joint convexity of D_α for probability distributions (A_1, B_1) and (A_2, B_2) and a parameter $0 < \delta < 1$ [38]:

$$\begin{aligned} D_\alpha((1 - \delta)A_1 + \delta A_2 \parallel (1 - \delta)B_1 + \delta B_2) \\ \leq (1 - \delta)D_\alpha(A_1 \parallel B_1) + \delta D_\alpha(A_2 \parallel B_2). \end{aligned} \tag{B13}$$

Setting $B_1 = B_2 = A_2 = q$ and $A_1 = p$, we find that

$$\begin{aligned} D_\alpha((1 - \delta)p + \delta q \parallel q) &\leq (1 - \delta)D_\alpha(p \parallel q) \\ &< D_\alpha(p \parallel q), \end{aligned} \tag{B14}$$

where the first inequality follows because $D_\alpha(q \parallel q) = 0$ [see (A13)] and the second strict inequality follows because $D_\alpha(p \parallel q) > 0$ for $p \neq q$ [see (A12)].

For $\alpha > 1$, joint convexity does not hold, and so a different approach is required. Let us define $r := (1 - \delta)p + \delta q$. Then the statement of the lemma is equivalent to the following:

$$\sum_{i=1}^k r_i^\alpha q_i^{1-\alpha} < \sum_{i=1}^k p_i^\alpha q_i^{1-\alpha}. \tag{B15}$$

For $\alpha > 1$, $f(x) = x^\alpha$ is a convex function over all $x > 0$. Since $q_i^{1-\alpha} > 0$, the following function

$$Q_\alpha(p) := \sum_{i=1}^k p_i^\alpha q_i^{1-\alpha} \tag{B16}$$

is convex with respect to p . For $p \neq q$, $Q_\alpha(p) > 1$, which follows from strict positivity of the Rényi divergence $D_\alpha(p \parallel q)$. Note that $Q_\alpha(q) = \sum_{i=1}^k q_i = 1$. Thus,

$$\begin{aligned} Q_\alpha(r) &\leq (1 - \delta)Q_\alpha(p) + \delta Q_\alpha(q) \\ &= Q_\alpha(p) - \delta(Q_\alpha(p) - Q_\alpha(q)) \\ &< Q_\alpha(p), \end{aligned} \tag{B17}$$

where the first inequality follows from convexity and the second inequality follows because $Q_\alpha(p) > Q_\alpha(q)$. Thus, we have established (B15).

For $\alpha < 0$, if p is not full rank, then $D_\alpha(p \parallel q) = \infty$. However, q is full rank, implying that $D_\alpha((1 - \delta)p + \delta q \parallel q)$ is finite. Thus, the desired inequality holds. If p is full rank, we use a similar approach to that used in the case $\alpha > 1$ by appealing to the identity in (A5), thus concluding the proof. \blacksquare

Lemma 5. Let r and s be probability distributions over the same alphabet. Then

$$\frac{1}{2} \|r - s\|_1 = \sum_{i:r_i > s_i} r_i - s_i. \tag{B18}$$

Proof. Consider that

$$1 = \sum_{i:r_i > s_i} r_i + \sum_{i:r_i \leq s_i} r_i = \sum_{i:r_i > s_i} s_i + \sum_{i:r_i \leq s_i} s_i. \tag{B19}$$

This implies that

$$\sum_{i:r_i > s_i} r_i - s_i = \sum_{i:r_i \leq s_i} s_i - r_i. \tag{B20}$$

Then we find that

$$\|r - s\|_1 = \sum_i |r_i - s_i| \tag{B21}$$

$$= \left[\sum_{i:r_i > s_i} r_i - s_i \right] + \left[\sum_{i:r_i \leq s_i} s_i - r_i \right] \tag{B22}$$

$$= 2 \sum_{i:r_i > s_i} r_i - s_i. \tag{B23}$$

This concludes the proof. ■

Definition 1 (Reversal channel). As is well known, a classical channel E is defined in terms of a conditional probability distribution. This implies that, given an input distribution $p(x)$ and a channel $E(y|x)$, the output distribution $p'(y)$ is given by

$$p'(y) = \sum_x E(y|x)p(x). \tag{B24}$$

A reversal channel for the classical channel E on p can be defined as follows. From the Bayes theorem, we know that

$$E(y|x)p(x) = E'(x|y)p'(y). \tag{B25}$$

Summing both sides of (B25) over x and noticing that $\sum_x E'(x|y) = 1$ is the normalization condition, we recover (B24). Similarly, summing both sides of (B25) over y , we see that

$$p(x) = \sum_y E'(x|y)p'(y), \tag{B26}$$

and thus a reversal channel E' can be defined as

$$E'(x|y) = \frac{E(y|x)p(x)}{p'(y)}. \tag{B27}$$

Equations (B28), (B29), and (B30) of the following lemma were established in [18], and the proof given below follows the proof given there closely.

Lemma 6 (Conversion into rational entries). Let $q = \{q_i\}_{i=1}^k$ be an ordered (descending) probability distribution of full rank. Then, for all $\varepsilon > 0$, there exists a probability distribution \tilde{q} such that

$$\frac{1}{2} \|q - \tilde{q}\|_1 \leq \varepsilon, \tag{B28}$$

where $\tilde{q} = \{d_i/N\}_i$ for a set $\{d_i\}_{i=1}^k$ of natural numbers, with $N := \sum_{i=1}^k d_i$. Additionally, there exists a classical channel E such that

$$\tilde{q} = E(q), \tag{B29}$$

and for all other probability distributions p , the following inequalities hold:

$$\frac{1}{2} \|p - E(p)\|_1 \leq O(\sqrt{\varepsilon}), \tag{B30}$$

$$\frac{1}{2} \|p - R(p)\|_1 \leq O(\sqrt{\varepsilon}), \tag{B31}$$

where R is the reversal channel for E on q , as given in Definition 1.

Proof. To begin with, if q is already rational, then the lemma trivially holds with $\tilde{q} = q$ and E and R set to the identity channel.

So we move on to the nontrivial case in which q is not rational. We first construct a probability distribution \tilde{q} that satisfies (B28). Then we construct a classical channel E that satisfies (B29) and (B30). We finally construct the reversal channel R that satisfies (B31).

Fix $\varepsilon > 0$. Since q is decreasing, it follows that $\min_i q_i = q_k$. Pick

$$N \geq \max \left\{ \left(\frac{k+1}{q_k} \right)^2, \frac{k}{\varepsilon}, 4 \right\}. \tag{B32}$$

Then N is sufficiently large such that the following inequality holds:

$$q_k > \frac{k}{\sqrt{N}} > \frac{k}{N}. \tag{B33}$$

We now define \tilde{q} . For $i \in \{1, \dots, k-1\}$, set

$$d_i := \lceil q_i N \rceil, \tag{B34}$$

$$\tilde{q}_i := \frac{d_i}{N}, \tag{B35}$$

and

$$d_k := N - \sum_{i=1}^{k-1} d_i, \tag{B36}$$

$$\tilde{q}_k = \frac{d_k}{N} = 1 - \sum_{i=1}^{k-1} \tilde{q}_i. \tag{B37}$$

Note that $\tilde{q}_i \geq q_i$ for $i \in \{1, \dots, k-1\}$ and $\tilde{q}_k \leq q_k$ because both distributions are normalized. In fact, we can conclude that $\tilde{q}_k < q_k$ from the assumption that q is not rational. Also note that

$$\begin{aligned} 1 - \tilde{q}_k &= \sum_{i=1}^{k-1} \tilde{q}_i \\ &= \sum_{i=1}^{k-1} \frac{\lceil q_i N \rceil}{N} \\ &\leq \sum_{i=1}^{k-1} \frac{q_i N + 1}{N} \\ &= \left(\sum_{i=1}^{k-1} \frac{q_i N}{N} \right) + \frac{k-1}{N} \\ &\leq 1 - q_k + \frac{k}{N}, \end{aligned} \tag{B38}$$

where the first equality follows from (B37) and the second equality follows from (B35). Thus,

$$\tilde{q}_k \geq q_k - \frac{k}{N} > 0, \tag{B39}$$

which follows from (B33) by the choice of N . Thus \tilde{q} is a legitimate probability distribution of full rank. Now applying Lemma 5, we find that

$$\begin{aligned} \frac{1}{2} \|q - \tilde{q}\|_1 &= \sum_{i:q_i > \tilde{q}_i} q_i - \tilde{q}_i, \\ &= q_k - \tilde{q}_k, \\ &\leq \frac{k}{N} \leq \varepsilon, \end{aligned} \tag{B40}$$

where the last inequalities are due to (B39) and (B32). Thus, it follows that (B28) is satisfied.

Now we construct a classical channel that takes q to \tilde{q} . Such a channel must slightly increase the probabilities q_i for all $i \in \{1, \dots, k-1\}$ while reducing q_k . Recall that classical channels are characterized by conditional probabilities $P(j|i)$ satisfying

$$P(j|i) \geq 0 \quad \forall i, j, \quad \sum_{j=1}^k P(j|i) = 1. \tag{B41}$$

Set $\mathcal{I} := \{1, 2, \dots, k-1\}$, which is the set of indices for which $\tilde{q}_i \geq q_i$. For $i \in \mathcal{I}$, let

$$\Delta_i := \tilde{q}_i - q_i, \tag{B42}$$

$$\Delta := \sum_{i \in \mathcal{I}} \Delta_i. \tag{B43}$$

Note that $\tilde{q}_k = q_k - \Delta$.

Consider a classical channel E defined as follows:

$$P(j|i) := \begin{cases} \delta_{ij}, & i, j \in \mathcal{I}, \\ 0, & i \in \mathcal{I}, j = k, \\ \frac{\Delta_j}{q_k}, & i = k, j \in \mathcal{I}, \\ 1 - \frac{\Delta}{q_k}, & i = j = k. \end{cases} \tag{B44}$$

It is clear that the non-negativity condition in (B41) holds, and we now prove that (i) the normalization condition holds, and (ii) $E(q) = \tilde{q}$. We begin with the normalization condition. For $i \in \mathcal{I}$, we have that

$$\begin{aligned} \sum_{j=1}^k P(j|i) &= P(k|i) + P(i|i) + \sum_{j \in \mathcal{I}, j \neq i} P(j|i) \\ &= 0 + 1 + 0 \\ &= 1. \end{aligned} \tag{B45}$$

For $i = k$, we have that

$$\begin{aligned} \sum_{j=1}^k P(j|k) &= P(k|k) + \sum_{j \in \mathcal{I}} P(j|k) \\ &= 1 - \frac{\Delta}{q_k} + \sum_{j \in \mathcal{I}} \frac{\Delta_j}{q_k} \\ &= 1. \end{aligned} \tag{B46}$$

Thus, the following equality holds for all $i \in \{1, \dots, k\}$:

$$\sum_{j=1}^k P(j|i) = 1, \tag{B47}$$

which corresponds to the preservation of normalization.

Next, we show that $E(q) = \tilde{q}$. For $j \in \mathcal{I}$, we have that

$$\begin{aligned} [E(q)]_j &= \sum_{i=1}^k P(j|i)q_i \\ &= P(j|j)q_j + P(j|k)q_k + \sum_{i \in \mathcal{I}, i \neq j} P(j|i)q_i \\ &= q_j + \frac{\Delta_j}{q_k}q_k + 0 \\ &= q_j + \Delta_j \\ &= \tilde{q}_j. \end{aligned} \tag{B48}$$

For $j = k$, we have that

$$\begin{aligned} [E(q)]_k &= \sum_{i=1}^k P(k|i)q_i \\ &= P(k|k)q_k + \sum_{i \in \mathcal{I}} P(k|i)q_i \\ &= \left(1 - \frac{\Delta}{q_k}\right)q_k + 0 \\ &= q_k - \Delta \\ &= \tilde{q}_k. \end{aligned} \tag{B49}$$

Thus, we have proven that $\tilde{q} = E(q)$.

We now need to show that, for all other distributions p , the following inequality holds:

$$\|p - E(p)\|_1 \leq O(\sqrt{\varepsilon}). \tag{B50}$$

Let $\tilde{p} := E(p)$. If $\tilde{p} = p$, then the desired inequality trivially holds. Otherwise, consider that $\tilde{p}_j \geq p_j$ for all $j \neq k$ and $p_k > \tilde{p}_k$, due to the form of the classical channel E in (B44). Indeed, this follows because, for $j \neq k$, we have that

$$\begin{aligned} \tilde{p}_j &= \sum_{i=1}^k P(j|i)p_i \\ &= P(j|j)p_j + P(j|k)p_k + \sum_{i \in \mathcal{I}, i \neq j} P(j|i)p_i \\ &= p_j + \Delta_j p_k / q_k \\ &\geq p_j, \end{aligned} \tag{B51}$$

while for $j = k$ we have that

$$\begin{aligned} \tilde{p}_k &= \sum_{i=1}^k P(k|i)p_i \\ &= P(k|k)p_k + \sum_{i \in \mathcal{I}} P(k|i)p_i \\ &= \left(1 - \frac{\Delta}{q_k}\right)p_k < p_k. \end{aligned} \tag{B52}$$

Now consider that

$$\begin{aligned} \frac{1}{2} \|p - \tilde{p}\|_1 &= p_k - \tilde{p}_k \\ &= p_k - \left(1 - \frac{\Delta}{q_k}\right) p_k \\ &= \left(\frac{\Delta}{q_k}\right) p_k \\ &\leq \frac{\Delta}{q_k} \\ &\leq \frac{k}{Nq_k} \\ &\leq \frac{1}{\sqrt{N}} \\ &\leq \sqrt{\frac{\varepsilon}{k}}, \end{aligned} \tag{B53}$$

where the first equality follows from Lemma 5 and the reasoning above, the first inequality follows because $p_k \leq 1$, the second inequality follows because $\Delta = q_k - \tilde{q}_k \leq \frac{k}{N}$, the third inequality follows because $q_k \geq \frac{k}{\sqrt{N}}$, and the last inequality follows because $\sqrt{\frac{k}{N}} \leq \sqrt{\varepsilon}$. Thus, for all other distributions p , the following inequality holds: $\|p - E(p)\|_1 \leq O(\sqrt{\varepsilon})$.

By applying Definition 1, the reversal channel R for E on q is defined as follows:

$$R(i|j) := \frac{P(j|i)q_i}{\tilde{q}_j}. \tag{B54}$$

For $i, j \in \mathcal{I}$, this becomes

$$R(i|j) = \frac{\delta_{i,j}q_i}{\tilde{q}_j}. \tag{B55}$$

For $i \in \mathcal{I}$ and $j = k$, we have

$$R(i|j) = 0. \tag{B56}$$

For $i = k$ and $j \in \mathcal{I}$, we have

$$R(i|j) = \frac{\Delta_j q_k}{q_k \tilde{q}_j} = \frac{\Delta_j}{\tilde{q}_j}. \tag{B57}$$

For $i = j = k$, we have

$$R(i|j) = \left(1 - \frac{\Delta}{q_k}\right) \frac{q_k}{\tilde{q}_k} = \frac{q_k}{\tilde{q}_k} - \frac{\Delta}{q_k \tilde{q}_k} = \frac{q_k - \Delta}{\tilde{q}_k} = 1. \tag{B58}$$

Summarizing all of this, the reversal channel $R(i|j)$ can be written as

$$R(i|j) = \begin{cases} \frac{\delta_{i,j}q_i}{\tilde{q}_j}, & i, j \in \mathcal{I}, \\ 0, & i \in \mathcal{I}, j = k, \\ \frac{\Delta_j}{\tilde{q}_j}, & i = k, j \in \mathcal{I}, \\ 1, & i = j = k. \end{cases} \tag{B59}$$

The action of this reversal channel on an arbitrary probability distribution p , leading to an output distribution Rp , is as

follows. For $i \neq k$, we have that

$$\begin{aligned} (Rp)_i &= \sum_{j=1}^k R(i|j)p_j \\ &= R(i|k)p_k + R(i|i)p_i + \sum_{j \in \mathcal{I}, j \neq i} R(i|j)p_j \\ &= 0 + \frac{q_i p_i}{\tilde{q}_i} + 0 \\ &= \frac{q_i p_i}{\tilde{q}_i} \\ &\leq p_i, \end{aligned} \tag{B60}$$

and for $i = k$, we have that

$$\begin{aligned} (Rp)_k &= \sum_{j=1}^k R(k|j)p_j \\ &= R(k|k)p_k + \sum_{j \in \mathcal{I}} R(k|j)p_j \\ &= p_k + \sum_{j \in \mathcal{I}} \frac{\Delta_j}{\tilde{q}_j} p_j \\ &> p_k. \end{aligned} \tag{B61}$$

Then we establish the following bound:

$$\begin{aligned} \frac{1}{2} \|Rp - p\|_1 &= (Rp)_k - p_k \\ &= \sum_{j \in \mathcal{I}} \frac{\Delta_j}{\tilde{q}_j} p_j \\ &\leq \frac{1}{\tilde{q}_k} \sum_{j \in \mathcal{I}} \Delta_j \\ &= \frac{\Delta}{\tilde{q}_k} \\ &\leq \frac{k}{N\tilde{q}_k} \\ &\leq \frac{2}{\sqrt{N}} \\ &\leq 2\sqrt{\frac{\varepsilon}{k}}. \end{aligned} \tag{B62}$$

The first equality follows because $(Rp)_k > p_k$ and $(Rp)_i \leq p_i$ for $i \neq k$, and by applying Lemma 5. The second equality follows from (B61). The first inequality follows because $p_j \leq 1$ and $\tilde{q}_j \geq \tilde{q}_k$ for all $j \in \mathcal{I}$. The third equality follows from (B43). The second inequality follows because $\Delta \leq k/N$. The third inequality follows from applying (B39) and (B33) to conclude that

$$\tilde{q}_k \geq q_k - \frac{k}{N} \geq \frac{k}{\sqrt{N}} - \frac{k}{N} \geq \frac{k}{2\sqrt{N}}, \tag{B63}$$

with the last inequality in (B63) following because

$$\frac{1}{\sqrt{N}} - \frac{1}{N} = \frac{\sqrt{N} - 1}{N} \geq \frac{\frac{1}{2}\sqrt{N}}{N} = \frac{1}{2\sqrt{N}} \tag{B64}$$

when $N \geq 4$, which is a consequence of (B32). The final inequality follows because $\sqrt{\frac{k}{N}} \leq \sqrt{\varepsilon}$. ■

Remark 1. For a full rank distribution p , the distribution $E(p)$ as defined above is full rank. The reasoning is as follows:

(i) The channel increases all entries except the last entry. Thus, all entries except the last entry in $E(p)$ are strictly greater than zero because p is full rank.

(ii) The last entry in $E(p)$ is $(1 - \frac{\Delta}{q_k})p_k$. Since $\Delta = q_k - \bar{q}_k$, by the choice of N , we see that this entry is strictly greater than zero as well [see (B39)].

Thus, the distribution $E(p)$ is full rank for a full rank p .

Lemma 7 (Splitting of channel [18]). Suppose that a channel, for some fixed input probability distribution $u = (u_1, \dots, u_\ell, 0, \dots, 0)$, where $u_1, \dots, u_\ell > 0$, outputs the following probability distribution: $u' = (u'_1, \dots, u'_\ell, 0, \dots, 0)$. Moreover, suppose that $\Lambda(w) = w$ holds for some full rank distribution w . Then $\Lambda = \Lambda_1 \oplus \Lambda_2$, where Λ_1 acts on the first ℓ elements and outputs to the first ℓ elements, and Λ_2 acts on the remaining $n - \ell$ elements and outputs to the remaining $n - \ell$ elements.

Proof. Let \mathcal{S} denote the set consisting of the first ℓ letters of the input alphabet (those for which $u_1, \dots, u_\ell > 0$). Consider the joint probability distribution induced by the input distribution w and the channel Λ . Let X be an indicator random variable, equal to 0 if the channel input is in \mathcal{S} and equal to 1 if the channel input is in \mathcal{S}^c . Similarly, let Y be an indicator random variable for \mathcal{S} in the same way for the channel output. Since the channel preserves w , it follows that

$$P(Y = 0) = P(X = 0), \tag{B65}$$

$$P(Y = 1) = P(X = 1). \tag{B66}$$

Moreover, since the channel transforms the input probability distribution u to u' , it follows that

$$\sum_{j \in \mathcal{S}} \Lambda_{j|i} = 1 \quad \forall i \in \mathcal{S}. \tag{B67}$$

To see this, consider that

$$u'_j = \sum_i \Lambda_{j|i} u_i = \sum_{i \in \mathcal{S}} \Lambda_{j|i} u_i, \tag{B68}$$

where the second equality follows because $u_i = 0$ if $i \in \mathcal{S}^c$. Now summing over $j \in \mathcal{S}$, we find that

$$\sum_{j \in \mathcal{S}} u'_j = 1, \tag{B69}$$

again because \mathcal{S}^c has zero probability mass for the output probability distribution u' . Then we conclude that

$$1 = \sum_{j \in \mathcal{S}} \sum_{i \in \mathcal{S}} \Lambda_{j|i} u_i = \sum_{i \in \mathcal{S}} u_i \sum_{j \in \mathcal{S}} \Lambda_{j|i}. \tag{B70}$$

From this condition, we can conclude that (B67) holds. To prove this claim, we use the method of contradiction. Suppose that it is not true, i.e., that there exists some $i \in \mathcal{S}$ such that $\sum_{j \in \mathcal{S}} \Lambda_{j|i} < 1$. Then substituting back into the right-hand side of (B70) and using the fact that $0 < u_i < 1$ for all $i \in \mathcal{S}$, it follows that the sum in (B70) is strictly less than 1, thus giving a contradiction.

We now claim that the condition in (B67) implies that $P(Y = 0|X = 0) = 1$. To see this, consider that

$$P(Y = 0|X = 0) = \frac{P(Y = 0, X = 0)}{P(X = 0)} \tag{B71}$$

$$= \frac{\sum_{i,j \in \mathcal{S}} \Lambda_{j|i} w_i}{\sum_{i \in \mathcal{S}} w_i} \tag{B72}$$

$$= \frac{\sum_{i \in \mathcal{S}} w_i \sum_{j \in \mathcal{S}} \Lambda_{j|i}}{\sum_{i \in \mathcal{S}} w_i} \tag{B73}$$

$$= \frac{\sum_{i \in \mathcal{S}} w_i}{\sum_{i \in \mathcal{S}} w_i} \tag{B74}$$

$$= 1. \tag{B75}$$

Thus, we conclude that

$$P(Y = 0) = P(Y = 0|X = 0)P(X = 0) + P(Y = 0|X = 1)P(X = 1), \tag{B76}$$

$$= P(Y = 0) + P(Y = 0|X = 1)P(X = 1), \tag{B77}$$

implying that

$$P(Y = 0|X = 1)P(X = 1) = 0. \tag{B78}$$

Since w is full rank, it follows that $P(X = 1) > 0$. We then conclude that

$$P(Y = 0|X = 1) = 0, \tag{B79}$$

from which we conclude that

$$P(Y = 1|X = 1) = 1. \tag{B80}$$

We can then rewrite this as follows:

$$1 = P(Y = 1|X = 1) = \frac{P(Y = 1, X = 1)}{P(X = 1)} \tag{B81}$$

$$= \frac{\sum_{i,j \in \mathcal{S}^c} \Lambda_{j|i} w_i}{\sum_{i \in \mathcal{S}^c} w_i} \tag{B82}$$

$$= \frac{\sum_{i \in \mathcal{S}^c} w_i \sum_{j \in \mathcal{S}^c} \Lambda_{j|i}}{\sum_{i' \in \mathcal{S}^c} w_{i'}} \tag{B83}$$

$$= \sum_{i \in \mathcal{S}^c} w'_i \sum_{j \in \mathcal{S}^c} \Lambda_{j|i}, \tag{B84}$$

where w'_i denotes the following probability distribution on \mathcal{S}^c :

$$w'_i := w_i / \sum_{i' \in \mathcal{S}^c} w_{i'}. \tag{B85}$$

Now following the same reasoning used to arrive at (B67), we conclude that

$$\sum_{j \in \mathcal{S}^c} \Lambda_{j|i} = 1 \quad \forall i \in \mathcal{S}^c. \tag{B86}$$

Combining (B67) and (B86), we conclude that first-group elements are always mapped to the first group and similarly for the second group. So the channel can be written as a direct sum of two channels as indicated. ■

Lemma 8 (Continuity [38]). The relative entropy $D(p||q)$ is continuous in both arguments p and q , when q has full rank.

Lemma 9 (Inclusion of support [41]). Let ρ_{AB} be a density operator acting on $\mathcal{H}_A \otimes \mathcal{H}_B$. Then

$$\text{supp}(\rho_{AB}) \subseteq \text{supp}(\rho_A) \otimes \text{supp}(\rho_B). \quad (\text{B87})$$

Proof. First let us suppose that ρ_{AB} is a pure state. Then $\rho_{AB} = |\psi\rangle\langle\psi|$. Let $|\psi\rangle = \sum_{z \in \mathcal{Z}} \alpha_z |\phi^z\rangle \otimes |\psi^z\rangle$ be the Schmidt decomposition of $|\psi\rangle$. Then

$$\begin{aligned} \text{supp}(\rho_{AB}) &= \{|\psi\rangle\} \\ &\subseteq \text{span}\{|\phi^z\rangle_{z \in \mathcal{Z}}\} \otimes \text{span}\{|\psi^z\rangle_{z \in \mathcal{Z}}\}. \end{aligned} \quad (\text{B88})$$

Since $\text{span}\{|\phi^z\rangle_{z \in \mathcal{Z}}\} = \text{supp}(\rho_A)$ and $\text{span}\{|\psi^z\rangle_{z \in \mathcal{Z}}\} = \text{supp}(\rho_B)$, we see that the lemma holds in this case.

For mixed states, let $\rho_{AB} = \sum_{x \in \mathcal{X}} \rho_{AB}^x$ be a decomposition into pure states for ρ_{AB} . Then

$$\begin{aligned} \text{supp}(\rho_{AB}) &= \text{span}\left\{\bigcup_{x \in \mathcal{X}} \text{supp}(\rho_{AB}^x)\right\} \\ &\subseteq \text{span}\left\{\bigcup_{x \in \mathcal{X}} \text{supp}(\rho_A^x) \otimes \text{supp}(\rho_B^x)\right\} \\ &\subseteq \left(\text{span}\left\{\bigcup_{x \in \mathcal{X}} \text{supp}(\rho_A^x)\right\}\right) \\ &\quad \otimes \left(\text{span}\left\{\bigcup_{x \in \mathcal{X}} \text{supp}(\rho_B^x)\right\}\right) \\ &= \text{supp}(\rho_A) \otimes \text{supp}(\rho_B). \end{aligned} \quad (\text{B89})$$

This concludes the proof. \blacksquare

Lemma 10 (Superadditivity of relative entropy [42]). Let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ be a bipartite Hilbert space, and let ρ_{AB} , σ_A , and σ_B be density operators. Then

$$\begin{aligned} D(\rho_{AB} \| \sigma_A \otimes \sigma_B) \\ = D(\rho_{AB} \| \rho_A \otimes \rho_B) + D(\rho_A \| \sigma_A) + D(\rho_B \| \sigma_B). \end{aligned} \quad (\text{B90})$$

Thus,

$$D(\rho_{AB} \| \sigma_A \otimes \sigma_B) \geq D(\rho_A \| \sigma_A) + D(\rho_B \| \sigma_B), \quad (\text{B91})$$

and the equality holds if and only if $\rho_{AB} = \rho_A \otimes \rho_B$.

Proof. Using the definition of quantum relative entropy, we find that

$$\begin{aligned} D(\rho_{AB} \| \sigma_A \otimes \sigma_B) \\ = \text{Tr}[\rho_{AB}(\ln \rho_{AB} - \ln \sigma_A \otimes \sigma_B)] \\ = \text{Tr}[\rho_{AB}(\ln \rho_{AB} - \ln \rho_A \otimes \rho_B)] \\ \quad + \text{Tr}[\rho_{AB}(\ln \rho_A \otimes \rho_B - \ln \sigma_A \otimes \sigma_B)] \\ = D(\rho_{AB} \| \rho_A \otimes \rho_B) + D(\rho_A \otimes \rho_B \| \sigma_A \otimes \sigma_B) \\ = D(\rho_{AB} \| \rho_A \otimes \rho_B) + D(\rho_A \| \sigma_A) + D(\rho_B \| \sigma_B). \end{aligned} \quad (\text{B92})$$

Since $D(\rho_{AB} \| \rho_A \otimes \rho_B) \geq 0$ [see (A25)], we conclude that $D(\rho_{AB} \| \sigma_A \otimes \sigma_B) \geq D(\rho_A \| \sigma_A) + D(\rho_B \| \sigma_B)$. The conclusion about equality holding is a direct consequence of $D(\rho_{AB} \| \rho_A \otimes \rho_B) = 0$ if and only if $\rho_{AB} = \rho_A \otimes \rho_B$. \blacksquare

Lemma 11 (Superadditivity of D_0). Let $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$ be a bipartite Hilbert space, and let ρ_{AB} , σ_A , and σ_B be density

operators. Then

$$D_0(\rho_{AB} \| \sigma_A \otimes \sigma_B) \geq D_0(\rho_A \| \sigma_A) + D_0(\rho_B \| \sigma_B). \quad (\text{B93})$$

Proof. From Lemma 9, we see that $\text{supp}(\rho_{AB}) \subseteq \text{supp}(\rho_A) \otimes \text{supp}(\rho_B)$ and thus the following operator inequality holds:

$$\Pi_{\rho_{AB}} \leq \Pi_{\rho_A} \otimes \Pi_{\rho_B}, \quad (\text{B94})$$

where Π_ω denotes the projection onto the support of the state ω . Then,

$$\begin{aligned} D_0(\rho_{AB} \| \sigma_A \otimes \sigma_B) \\ = -\ln(\text{Tr}[\Pi_{\rho_{AB}}(\sigma_A \otimes \sigma_B)]) \\ \geq -\ln(\text{Tr}[(\Pi_{\rho_A} \otimes \Pi_{\rho_B})(\sigma_A \otimes \sigma_B)]) \\ = -\ln(\text{Tr}[(\Pi_{\rho_A} \sigma_A \otimes \Pi_{\rho_B} \sigma_B)]) \\ = -\ln(\text{Tr}[\Pi_{\rho_A} \sigma_A]) - \ln(\text{Tr}[\Pi_{\rho_B} \sigma_B]) \\ = D_0(\rho_A \| \sigma_A) + D_0(\rho_B \| \sigma_B). \end{aligned} \quad (\text{B95})$$

This concludes the proof. \blacksquare

Lemmas 11 and 12 are results based on quantum divergences. However, any result that holds for the quantum case also holds for the classical case. This is because we can plug in commuting quantum states as a special case, and these states are quasiclassical.

Lemma 12 (Müller [23]). Let $p_A, p'_A \in \mathbb{R}^k$ be probability distributions with $p_A^\downarrow \neq p'_A^\downarrow$. Then there exists a probability distribution q_B and an extension p'_{AB} with marginals p'_A and q_B such that

$$p_A \otimes q_B \succ p'_{AB} \quad (\text{B96})$$

if and only if $H_0(p_A) \leq H_0(p'_A)$ and $H(p_A) < H(p'_A)$. Moreover, if these inequalities are satisfied, then for all $\varepsilon > 0$ we can choose system B and p'_{AB} such that

$$D(p'_{AB} \| p'_A \otimes q_B) < \varepsilon. \quad (\text{B97})$$

APPENDIX C: PROOF OF THEOREM 1

Proof. Since q and q' have rational entries, without loss of generality, we pick $q = (\frac{d_1}{N}, \dots, \frac{d_k}{N})$ and $q' = (\frac{d'_1}{N}, \dots, \frac{d'_k}{N})$ for sets $\{d_i\}_i$ and $\{d'_i\}_i$ of natural numbers, such that

$$\sum_{i=1}^k d_i = \sum_{i=1}^k d'_i = N, \quad (\text{C1})$$

respectively.

We begin with statement 1 implies statement 2; i.e., we suppose that $D(p \| q) > D(p' \| q')$ and $D_0(p \| q) \geq D_0(p' \| q')$, and we prove the existence of a classical channel Λ , a probability distribution r , and a joint distribution t' that satisfy the conditions of statement 2.

Using Lemma 3 and statement 1, we conclude that

$$\begin{aligned} D(\Gamma_d(p) \| \eta_N) &> D(\Gamma_{d'}(p') \| \eta_N), \\ D_0(\Gamma_d(p) \| \eta_N) &\geq D_0(\Gamma_{d'}(p') \| \eta_N). \end{aligned} \quad (\text{C2})$$

Define $\tilde{p} \equiv \Gamma_d(p)$ and $\tilde{p}' \equiv \Gamma_{d'}(p')$. Thus,

$$\begin{aligned} D(\tilde{p} \parallel \eta_N) &> D(\tilde{p}' \parallel \eta_N), \\ D_0(\tilde{p} \parallel \eta_N) &\geq D_0(\tilde{p}' \parallel \eta_N). \end{aligned} \tag{C3}$$

Recalling (A20), the following equality holds for all probability distributions \tilde{p} and $\alpha \in \{0, 1\}$:

$$D_\alpha(\tilde{p} \parallel \eta_N) = \ln(N) - H_\alpha(\tilde{p}). \tag{C4}$$

Thus, from (C3) and (C4), we conclude that

$$\begin{aligned} H(\tilde{p}) &< H(\tilde{p}'), \\ H_0(\tilde{p}) &\leq H_0(\tilde{p}'). \end{aligned} \tag{C5}$$

Furthermore, we conclude that

$$\tilde{p}^\downarrow \neq \tilde{p}'^\downarrow, \tag{C6}$$

which follows from the structure of the channels Γ_d and $\Gamma_{d'}$, as well as the assumption that the relative spectra of the pairs (p, q) and (p', q') are different.

Using (C5) and (C6), we apply Lemma 12 to conclude that there exists a probability distribution r and an extension v' with marginals \tilde{p}' and r , such that

$$\begin{aligned} \tilde{p} \otimes r &> v', \\ D(v' \parallel \tilde{p}' \otimes r) &< \gamma \end{aligned} \tag{C7}$$

for all $\gamma > 0$. In other words, there exists a doubly stochastic channel Φ (see Lemma 1) such that

$$\Phi(\tilde{p} \otimes r) = v'. \tag{C8}$$

Using Lemma 7, we conclude that r can be considered, without loss of generality, to be of full rank. Concretely, if r is not full rank, then define $u = \tilde{p} \otimes r$, $u' = v'$, and $w = \eta_N \otimes \eta$. Using Lemma 7, Φ can be split into $\Phi_1 \oplus \Phi_2$. If r is of full rank, then $\Phi = \Phi_1$.

Finally consider the following channel:

$$\Lambda := (\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1 \circ (\Gamma_d \otimes \text{id}). \tag{C9}$$

Since this is a composition of classical channels, the overall map is a classical channel.

Proving 2a.

$$\begin{aligned} \Lambda(p \otimes r) &= [(\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1 \circ (\Gamma_d \otimes \text{id})](p \otimes r), \\ &= [(\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1](\tilde{p} \otimes r), \\ &= (\Gamma_{d'}^* \otimes \text{id})(v'). \end{aligned} \tag{C10}$$

We now define $t' = (\Gamma_{d'}^* \otimes \text{id})(v')$, and we observe that the marginals of t' are $\Gamma_{d'}^*(\tilde{p}') = p'$ and r .

Proving 2b.

We need to show that $\Lambda(q \otimes \eta) = q' \otimes \eta$. Consider that

$$\begin{aligned} \Lambda(q \otimes \eta) &= [(\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1 \circ (\Gamma_d \otimes \text{id})](q \otimes \eta), \\ &= [(\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1](\eta_N \otimes \eta), \\ &= (\Gamma_{d'}^* \otimes \text{id})(\eta_N \otimes \eta), \\ &= q' \otimes \eta. \end{aligned} \tag{C11}$$

where the second equality follows from Lemma 2. Thus, $\Lambda(q \otimes \eta) = q' \otimes \eta$.

Proving 2c.

$$\begin{aligned} \gamma &> D(v' \parallel \tilde{p}' \otimes r), \\ &\geq D(t' \parallel p' \otimes r), \end{aligned} \tag{C12}$$

where the first inequality follows from (C7) and the second inequality follows from the data-processing inequality in (A11), using the channel $\Gamma_{d'}^* \otimes \text{id}$. Thus, $D(t' \parallel p' \otimes r) \leq \gamma$.

This completes the proof that statement 1 implies statement 2.

We now establish the reverse implication, i.e., we suppose that there exists a classical channel Λ satisfying the conditions of statement 2. Let $\alpha \in \{0, 1\}$. Then,

$$\begin{aligned} D_\alpha(p \parallel q) + D_\alpha(r \parallel \eta) &= D_\alpha(p \otimes r \parallel q \otimes \eta) \\ &\geq D_\alpha(\Lambda(p \otimes r) \parallel \Lambda(q \otimes \eta)), \\ &= D_\alpha(t' \parallel q' \otimes \eta), \\ &\geq D_\alpha(p' \parallel q') + D_\alpha(r \parallel \eta), \end{aligned} \tag{C13}$$

where the first equality follows from the additivity of the Rényi divergence in (A14), the first inequality follows from the data-processing inequality in (A11), and the last inequality follows from Lemmas 10 and 11 for $\alpha = 1$ and 0, respectively.

Since the support of r is contained in the support of η , it follows that $D_\alpha(r \parallel \eta)$ is finite and can be subtracted, and we conclude that the following inequality holds for $\alpha \in \{0, 1\}$:

$$D_\alpha(p \parallel q) \geq D_\alpha(p' \parallel q'). \tag{C14}$$

To conclude that the inequality

$$D(p \parallel q) \geq D(p' \parallel q') \tag{C15}$$

is actually strict, consider that the equality $D(p \parallel q) = D(p' \parallel q')$ is equivalent to the equality in (C13). Saturating the last inequality of (C13) is then equivalent to $t' = p' \otimes r$. This in turn implies that the forward channel Λ realizes the following transformation:

$$\Lambda(p \otimes r) = p' \otimes r, \tag{C16}$$

$$\Lambda(q \otimes \eta) = q' \otimes \eta. \tag{C17}$$

Now defining $\tilde{p} = \Gamma_d(p)$ and $\tilde{p}' = \Gamma_{d'}(p')$ for N satisfying (C1), and recalling that $\eta_N = \Gamma_d(q)$ and $\eta_N = \Gamma_{d'}(q')$, we conclude that

$$[(\Gamma_{d'} \otimes \text{id}) \circ \Lambda \circ (\Gamma_d^* \otimes \text{id})](\tilde{p} \otimes r) = \tilde{p}' \otimes r, \tag{C18}$$

$$[(\Gamma_{d'} \otimes \text{id}) \circ \Lambda \circ (\Gamma_d^* \otimes \text{id})](\eta_N \otimes \eta) = \eta_N \otimes \eta. \tag{C19}$$

So this means that $(\Gamma_{d'} \otimes \text{id}) \circ \Lambda \circ (\Gamma_d^* \otimes \text{id})$ is a doubly stochastic channel, and in turn that $\tilde{p} \succ_T \tilde{p}'$ for $\tilde{p}^\downarrow \neq \tilde{p}'^\downarrow$ (the latter following from the assumption of differing relative spectra). By applying the main result of [29], it follows that $H(\tilde{p}) < H(\tilde{p}')$, which in turn is equivalent to $D(p \parallel q) > D(p' \parallel q')$ by applying (C4). This contradicts the possibility of equality in (C15), and so we conclude that only a strict inequality can hold. ■

APPENDIX D: PROOF OF THEOREM 2

Proof. Some aspects of the proof given below follow the proof of Ref. [18] (Theorem 17) closely.

We begin with statement 1 implies statement 2; i.e., we suppose that $D(p\|q) \geq D(p'\|q')$ and we prove the existence of a classical channel Λ , probability distributions p'_ε and r , and a joint distribution t'_ε that satisfy the conditions of statement 2.

As a first check, we can determine by means of an efficient algorithm (see, e.g., [14] or [11]) if the pair (p, q) relatively majorizes the pair (p', q') . If this is the case, then the claim trivially holds without catalysis because there exists a classical channel that takes p to p' and q to q' .

In the more general case of interest for this paper, it may not be the case that the pair (p, q) relatively majorizes the pair (p', q') . Furthermore, it is not necessarily the case that q and q' have rational entries. However, the set of probability distributions with rational entries is dense in the set of all distributions. Thus, using Lemma 6, we can always pick probability distributions q_d and $q'_{d'}$ (defined below) that are arbitrarily close to q and q' , respectively.

Without loss of generality, we pick $q_d = (\frac{d_1}{N}, \dots, \frac{d_k}{N})$ and $q'_{d'} = (\frac{d'_1}{N}, \dots, \frac{d'_k}{N})$ for integers $\{d_i\}$ and $\{d'_i\}$, such that $\sum_{i=1}^k d_i = \sum_{i=1}^k d'_i = N$, respectively, with N sufficiently large as needed for the application of Lemma 6.

Applying Lemma 6, we conclude that there exist classical channels E and E' such that $E(q) = q_d$ and $E'(q') = q'_{d'}$,

$$\begin{aligned} \frac{1}{2} \|q_d - q\|_1 &\leq \varepsilon_1, \\ \frac{1}{2} \|E(p) - p\|_1 &\leq O(\sqrt{\varepsilon_1}) \end{aligned} \quad (\text{D1})$$

for all probability distributions p , and

$$\begin{aligned} \frac{1}{2} \|q'_{d'} - q'\|_1 &\leq \varepsilon_2, \\ \frac{1}{2} \|E'(p') - p'\|_1 &\leq O(\sqrt{\varepsilon_2}) \end{aligned} \quad (\text{D2})$$

for all probability distributions p' .

Define the reversal channel of E' as E'^* (see Definition 1). Then

$$E'^*(q'_{d'}) = q' \quad (\text{D3})$$

and applying Lemma 6, it follows that

$$\frac{1}{2} \|E'^*(p') - p'\|_1 \leq O(\sqrt{\varepsilon_2}) \quad (\text{D4})$$

for all probability distributions p' . Now define

$$p'' = (1 - \delta)p' + \delta q' \quad (\text{D5})$$

for $0 < \delta < 1$. Note that

$$\frac{1}{2} \|p'' - p'\|_1 \leq \delta \quad (\text{D6})$$

and p'' is full rank. From Lemma 4 and statement 1, we conclude that

$$D(p\|q) > D(p''\|q'). \quad (\text{D7})$$

Lemma 8 states that $D(p\|q)$ is continuous in both arguments p and q when q is full rank. Thus it follows that

$$|D(p\|q) - D(E(p)\|q_d)| \leq f_1(\varepsilon_1) \quad (\text{D8})$$

for some function $f_1(\varepsilon_1)$ with the property that $\lim_{\varepsilon_1 \rightarrow 0} f_1(\varepsilon_1) = 0$. Similarly, it follows that

$$|D(p''\|q') - D(E'(p'')\|q'_{d'})| \leq f_2(\varepsilon_2) \quad (\text{D9})$$

for some function $f_2(\varepsilon_1)$ with the property that $\lim_{\varepsilon_1 \rightarrow 0} f_2(\varepsilon_1) = 0$. Thus, using (D7), (D8), and (D9), and taking ε_1 and ε_2 sufficiently small by taking N sufficiently large, so that

$$D(p\|q) - D(p''\|q') > f(\varepsilon_1) + f(\varepsilon_2), \quad (\text{D10})$$

we conclude that

$$D(E(p)\|q_d) > D(E'(p'')\|q'_{d'}). \quad (\text{D11})$$

Then using Lemma 3 and (D11), we conclude that

$$D(\Gamma_d(E(p))\|\eta_N) > D(\Gamma_{d'}(E'(p''))\|\eta_N). \quad (\text{D12})$$

Define

$$\tilde{E}(p) \equiv \Gamma_d(E(p)), \quad (\text{D13})$$

$$\tilde{E}'(p'') \equiv \Gamma_{d'}(E'(p'')). \quad (\text{D14})$$

Thus,

$$D(\tilde{E}(p)\|\eta_N) > D(\tilde{E}'(p'')\|\eta_N). \quad (\text{D15})$$

Applying (D15) and (A20), we conclude that

$$H(\tilde{E}(p)) < H(\tilde{E}'(p'')). \quad (\text{D16})$$

Since p'' is full rank, it follows from Remark 1 that $E'(p'')$ is full rank also, so that $\tilde{E}'(p'')$ is full rank as well. From the definition of H_0 (see Appendix A2), it follows that

$$H_0(\tilde{E}(p)) \leq H_0(\tilde{E}'(p'')). \quad (\text{D17})$$

Due to fact that (D16) holds, we can also conclude that

$$\tilde{E}(p)^\downarrow \neq \tilde{E}'(p'')^\downarrow. \quad (\text{D18})$$

For if we had $\tilde{E}(p)^\downarrow = \tilde{E}'(p'')^\downarrow$, then necessarily $H(\tilde{E}(p)) = H(\tilde{E}'(p''))$.

Since (D16), (D17), and (D18) hold, we can invoke Lemma 12 to conclude that there exists a probability distribution r and a joint distribution v'' with marginals $\tilde{E}'(p'')$ and r such that

$$\tilde{E}(p) \otimes r \succ v'', \quad (\text{D19})$$

$$D(v''\|\tilde{E}'(p'') \otimes r) < \gamma$$

for all $\gamma > 0$. Thus, there exists a doubly stochastic channel Φ (see Lemma 1) such that

$$\Phi(\tilde{E}(p) \otimes r) = v''. \quad (\text{D20})$$

Applying Lemma 7, it follows that r can be considered, without loss of generality, to be of full rank. Concretely, if r is not full rank, then define $u = \tilde{E}(p) \otimes r$, $u' = v''$, and $w = \eta_N \otimes \eta$. Using Lemma 7, Φ can be split into $\Phi_1 \oplus \Phi_2$. If r is of full rank, then $\Phi = \Phi_1$.

Finally, we define the following channel:

$$\Lambda = (E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1 \circ (\Gamma_d \otimes \text{id}) \circ (E \otimes \text{id}). \quad (\text{D21})$$

Since this is a composition of classical channels, the overall map is a classical channel.

Proving 2a and 2c.

$$\begin{aligned}
\Lambda(p \otimes r) &= [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1 \\
&\quad \circ (\Gamma_d \otimes \text{id}) \circ (E \otimes \text{id})](p \otimes r) \\
&= [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1 \circ (\Gamma_d \otimes \text{id})][E(p) \otimes r] \\
&= [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1][\tilde{E}(p) \otimes r] \\
&= [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id})]v''. \tag{D22}
\end{aligned}$$

We now define $t'_\varepsilon = [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id})]v''$. Observe that the marginals of t'_ε are $p'_\varepsilon = E'^*(\Gamma_{d'}^*(\tilde{E}'(p'')))$ and r .

Now consider the following chain of inequalities:

$$\begin{aligned}
\frac{1}{2}\|p'_\varepsilon - p'\|_1 &= \frac{1}{2}\|E'^*(\Gamma_{d'}^*(\tilde{E}'(p''))) - p'\|_1 \\
&\leq \frac{1}{2}\|E'^*(\Gamma_{d'}^*(\tilde{E}'(p''))) - \Gamma_{d'}^*(\tilde{E}'(p''))\|_1 \\
&\quad + \frac{1}{2}\|\Gamma_{d'}^*(\tilde{E}'(p'')) - p'\|_1 \\
&\leq O(\sqrt{\varepsilon_2}) + \frac{1}{2}\|\Gamma_{d'}^*(\Gamma_{d'}(E'(p''))) - p'\|_1 \\
&= O(\sqrt{\varepsilon_2}) + \frac{1}{2}\|E'(p'') - p'\|_1 \\
&\leq O(\sqrt{\varepsilon_2}) + \frac{1}{2}\|E'(p'') - p''\|_1 + \frac{1}{2}\|p'' - p'\|_1 \\
&\leq O(\sqrt{\varepsilon_2}) + O(\sqrt{\varepsilon_1}) + \delta, \tag{D23}
\end{aligned}$$

where the first inequality follows from the triangle inequality, the second inequality follows from (D4), the second equality follows from (A33), the third inequality follows from the triangle inequality, and the last inequality follows from (D2) and (D6).

Since ε_1 , ε_2 , and δ can be made arbitrarily small, we can set the right-hand side of (D23) to ε . Thus $\Lambda(p \otimes r) = t'_\varepsilon$ with marginals p'_ε and r , and the following inequality holds: $\frac{1}{2}\|p'_\varepsilon - p'\|_1 \leq \varepsilon$.

Proving 2b.

We need to show that $\Lambda(q \otimes \eta) = q' \otimes \eta$. Then,

$$\begin{aligned}
\Lambda(q \otimes \eta) &= [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1 \\
&\quad \circ (\Gamma_d \otimes \text{id}) \circ (E \otimes \text{id})](q \otimes \eta) \\
&= [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1 \circ (\Gamma_d \otimes \text{id})](q_d \otimes \eta) \\
&= [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id}) \circ \Phi_1](\eta_N \otimes \eta)
\end{aligned}$$

$$\begin{aligned}
&= [(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id})](\eta_N \otimes \eta) \\
&= [(E'^* \otimes \text{id})](q'_{d'} \otimes \eta) \\
&= q' \otimes \eta, \tag{D24}
\end{aligned}$$

where the third equality follows because Φ_1 is a doubly stochastic channel. Thus $\Lambda(q \otimes \eta) = q' \otimes \eta$.

Proving 2d.

$$\begin{aligned}
\gamma &> D(v'' \|\tilde{E}''(p') \otimes r) \\
&\geq D(t'_\varepsilon \| p'_\varepsilon \otimes r), \tag{D25}
\end{aligned}$$

where the first inequality follows from (D19) and the second inequality follows from the data-processing inequality (A11) using the channel $(E'^* \otimes \text{id}) \circ (\Gamma_{d'}^* \otimes \text{id})$. Thus, $D(t'_\varepsilon \| p'_\varepsilon \otimes r) \leq \gamma$. This completes the proof that statement 1 implies statement 2.

We now look at the reverse direction, i.e., fix $\varepsilon \in (0, 1)$ and $\gamma > 0$ and suppose that there are probability distributions r and p'_ε , a joint distribution t'_ε , and a classical channel Λ satisfying the stated conditions. Then consider that

$$\begin{aligned}
D(p\|q) + D(r\|\eta) &= D(p \otimes r \| q \otimes \eta) \\
&\geq D(\Lambda(p \otimes r) \| \Lambda(q \otimes \eta)) \\
&= D(t'_\varepsilon \| q' \otimes \eta) \\
&\geq D(p'_\varepsilon \| q') + D(r\|\eta), \tag{D26}
\end{aligned}$$

where the first equality follows from additivity of relative entropy (A14), the first inequality follows from the data-processing inequality (A11), and the last inequality follows from Lemma 10. Since η is full rank, $D(r\|s)$ is finite and can be subtracted, so that the following inequality holds:

$$D(p\|q) \geq D(p'_\varepsilon \| q'). \tag{D27}$$

Since this inequality holds for all $\varepsilon \in (0, 1)$, we can apply Lemma 8 and take the limit as $\varepsilon \rightarrow 0$ to conclude that

$$D(p\|q) \geq D(p' \| q'). \tag{D28}$$

This concludes the proof. ■

APPENDIX E: PROOF OF COROLLARY 1

Proof. Simply pick $q = q' = \eta$ and apply Theorem 2. Since Λ preserves the uniform distribution, it is a unital classical channel. ■

-
- [1] G. H. Hardy, D. E. Littlewood, and G. Pólya, *Inequalities*, Cambridge Mathematical Library (Cambridge University Press, Cambridge, 1952).
- [2] E. Chitambar and G. Gour, Quantum resource theories, *Rev. Mod. Phys.* **91**, 025001 (2019).
- [3] M. A. Nielsen, Conditions for a Class of Entanglement Transformations, *Phys. Rev. Lett.* **83**, 436 (1999).
- [4] S. Du, Z. Bai, and Y. Guo, Conditions for coherence transformations under incoherent operations, *Phys. Rev. A* **91**, 052120 (2015).
- [5] H. Zhu, Z. Ma, Z. Cao, S.-M. Fei, and V. Vedral, Operational one-to-one mapping between coherence and entanglement measures, *Phys. Rev. A* **96**, 032316 (2017).
- [6] A. Winter and D. Yang, Operational Resource Theory of Coherence, *Phys. Rev. Lett.* **116**, 120404 (2016).
- [7] M. Horodecki, P. Horodecki, and J. Oppenheim, Reversible transformations from pure to mixed states and the unique measure of information, *Phys. Rev. A* **67**, 062104 (2003).
- [8] E. Ruch and A. Mead, The principle of increasing mixing character and some of its consequences, *Theor. Chim. Acta* **41**, 95 (1976).
- [9] E. Ruch, R. Schraner, and T. H. Seligman, The mixing distance, *J. Chem. Phys.* **69**, 386 (1978).
- [10] E. Ruch, R. Schraner, and T. H. Seligman, Generalization of a theorem by Hardy, Littlewood, and Pólya, *J. Math. Anal. Appl.* **76**, 222 (1980).

- [11] J. M. Renes, Relative submajorization and its use in quantum resource theories, *J. Math. Phys.* **57**, 122202 (2016).
- [12] F. Buscemi and G. Gour, Quantum relative Lorenz curves, *Phys. Rev. A* **95**, 012110 (2017).
- [13] D. Blackwell, Equivalent comparisons of experiments, *Ann. Math. Stat.* **24**, 265 (1953).
- [14] G. Dahl, Matrix majorization, *Lin. Alg. Appl.* **288**, 53 (1999).
- [15] A. F. Veinott, Least d -majorized network flows with inventory and statistical applications, *Manag. Sci.* **17**, 547 (1971).
- [16] D. Jonathan and M. B. Plenio, Entanglement-Assisted Local Manipulation of Pure Quantum States, *Phys. Rev. Lett.* **83**, 3566 (1999).
- [17] S. Daftuar and M. Klimesh, Mathematical structure of entanglement catalysis, *Phys. Rev. A* **64**, 042314 (2001).
- [18] F. G. S. L. Brandão, M. Horodecki, N. Ng, J. Oppenheim, and S. Wehner, The second laws of quantum thermodynamics, *Proc. Natl. Acad. Sci. USA* **112**, 3275 (2015).
- [19] P. Boes, J. Eisert, R. Gallego, M. P. Müller, and H. Wilming, Von Neumann Entropy from Unitarity, *Phys. Rev. Lett.* **122**, 210402 (2019).
- [20] K. Bu, U. Singh, and J. Wu, Catalytic coherence transformations, *Phys. Rev. A* **93**, 042326 (2016).
- [21] S. Chen, X. Zhang, Y. Zhou, and Q. Zhao, One-shot coherence distillation with catalysts, *Phys. Rev. A* **100**, 042323 (2019).
- [22] The stated assumption about reuse of the catalyst is quite important, and violation of it could lead to problems in applications. To be clear, once the transformation from the distribution pair (p, q) to the distribution pair (p', q') is completed, it is necessary for an incoming distribution pair to be independent of (p, q) in order to reuse the catalyst. If it is not, the conditions of our theorem do not hold and the new transformation is not guaranteed to succeed. See [23] for further discussions.
- [23] M. P. Müller, Correlating Thermal Machines and the Second Law at the Nanoscale, *Phys. Rev. X* **8**, 041051 (2018).
- [24] K. Matsumoto, Reverse test and characterization of quantum relative entropy, [arXiv:1010.1030](https://arxiv.org/abs/1010.1030).
- [25] X. Wang and M. M. Wilde, Resource theory of asymmetric distinguishability, *Phys. Rev. Research* **1**, 033170 (2019).
- [26] F. Buscemi, D. Sutter, and M. Tomamichel, An information-theoretic treatment of quantum dichotomies, *Quantum* **3**, 209 (2019).
- [27] M. Lostaglio, The resource theory of quantum thermodynamics, Ph.D. thesis, Imperial College London, 2016, available at <https://doi.org/10.25560/43760>.
- [28] M. Horodecki and J. Oppenheim, Fundamental limitations for quantum and nanoscale thermodynamics, *Nat. Commun.* **4**, 2059 (2013).
- [29] M. Klimesh, Inequalities that collectively completely characterize the catalytic majorization relation, [arXiv:0709.3680](https://arxiv.org/abs/0709.3680).
- [30] S. Turgut, Catalytic transformations for bipartite pure states, *J. Phys. A* **40**, 12185 (2007).
- [31] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, Concentrating partial entanglement by local operations, *Phys. Rev. A* **53**, 2046 (1996).
- [32] F. G. S. L. Brandão, M. Horodecki, J. Oppenheim, J. M. Renes, and R. W. Spekkens, Resource Theory of Quantum States Out of Thermal Equilibrium, *Phys. Rev. Lett.* **111**, 250404 (2013).
- [33] K. Korzekwa, C. T. Chubb, and M. Tomamichel, Avoiding Irreversibility: Engineering Resonant Conversions of Quantum Resources, *Phys. Rev. Lett.* **122**, 110403 (2019).
- [34] P. M. Alberti and A. Uhlmann, A problem relating to positive linear maps on matrix algebras, *Rep. Math. Phys.* **18**, 163 (1980).
- [35] K. Matsumoto, 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](https://arxiv.org/abs/1409.5658).
- [36] G. Gour, D. Jennings, F. Buscemi, R. Duan, and I. Marvian, Quantum majorization and a complete set of entropic conditions for quantum thermodynamics, *Nat. Commun.* **9**, 5352 (2018).
- [37] A. Rényi, On measures of entropy and information, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics* (University of California Press, Berkeley, CA, 1961), pp. 547–561.
- [38] T. van Erven and P. Harremoës, Rényi divergence and Kullback-Leibler divergence, *IEEE Trans. Inf. Theor.* **60**, 3797 (2014).
- [39] D. Petz, Quasi-entropies for finite quantum systems, *Rep. Math. Phys.* **23**, 57 (1986).
- [40] A. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and its Applications* (Springer Science+Business Media, New York, 2011).
- [41] R. Renner, Security of quantum key distribution, Ph.D. thesis, ETH Zurich, 2005.
- [42] A. Capel, A. Lucia, and D. Pérez-García, Superadditivity of quantum relative entropy for general states, *IEEE Trans. Inf. Theor.* **64**, 4758 (2018).