Einstein-Podolsky-Rosen steering provides the advantage in entanglement-assisted subchannel discrimination with one-way measurements

Marco Piani[1] and John Watrous[2]

[1]*Institute for Quantum Computing & Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada*
[2]*Institute for Quantum Computing & School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada*

Steering is the entanglement-based quantum effect that embodies the "spooky action at a distance" disliked by Einstein and scrutinized by Einstein, Podolsky, and Rosen. Here we provide a necessary and sufficient characterization of steering, based on a quantum information processing task: the discrimination of branches in a quantum evolution, which we dub *subchannel discrimination.* We prove that, for any bipartite steerable state, there are instances of the quantum subchannel discrimination problem for which this state allows a correct discrimination with strictly higher probability than in absence of entanglement, even when measurements are restricted to local measurements aided by one-way communication. On the other hand, unsteerable states are useless in such conditions, even when entangled. We also prove that the above steering advantage can be exactly quantified in terms of the *steering robustness*, which is a natural measure of the steerability exhibited by the state.

PACS numbers: 03.67.Mn, 03.67.Bg, 03.65.Ud

Entanglement is a property of distributed quantum systems that does not have a classical counterpart [1]. On one hand, entanglement challenges our classical, everyday-life intuition about the physical world; on the other hand, it is the key element in many quantum information processing tasks [2]. The strongest feature exhibited by entangled systems is non-locality [3]. A weaker feature related to entanglement is *steering*: roughly speaking, it corresponds to the fact that one party can induce very different ensembles for the local state of the other party, beyond what is possible based only on a conceivable classical knowledge about the other party's "hidden state" [4, 5]. Steering embodies the "spooky action at a distance"—in the words of Einstein [6]—identified by Schroedinger [7], scrutinized by Einstein, Podolsky, and Rosen [8], and formally put on sound ground in [4, 5].

Not all entangled states are steerable, and not all steerable states exhibit nonlocality [4, 5], but states that exhibit steering allow for the verification of their entanglement in a *semi*-device independent way: there is no need to trust the devices used by the steering party, and the ability to determine the conditional states of the steered party is sufficient [4, 5, 9]. In general, besides its foundational interest, steering is interesting in practice in bipartite tasks, like quantum key distribution (QKD) [10], where it is convenient and/or appropriate to trust the devices of one of two parties, but not necessarily of the other party. For example, by exploiting steering it is possible to obtain key rates unachievable in a full device-independent approach [11], but still assuming less about the devices than in a standard QKD approach [12]. For these reasons, steering has recently attracted significant interest, both theoretically and experimentally [13–30], mostly directed to the verification of steering. Nonetheless, an answer to the question "What is steering useful for?" that applies to states that exhibit steering can arguably be considered limited [9, 12]. Furthermore, the quantification of steering has just started to be addressed [24].

In this Letter we fully characterize and quantify steering in an operational way that mirrors the asymmetric features of steering, and that breaks new ground in the investigation of the usefulness of steering. We prove that every steerable state is a resource in a quantum information task that we dub *subchannel discrimination*, in a practically relevant scenario where measurements can only be performed locally.

Subchannel discrimination is the identification of which branch of a quantum evolution a quantum system undergoes (see Fig. 1). It is well known that entanglement between a probe and an ancilla can help in discriminating different channels [31–41]. In [42] it was proven that *every* entangled state is useful in some instance of the subchannel discrimination problem. Ref. [43] raised and analyzed the question of whether such an advantage is preserved when joint measurements on the output probe and the ancilla are not possible. Here we prove that, when only local measurements coordinated by forward classical communication are possible, *every* steerable state remains useful, while non-steerable entangled states become useless. We further prove that this usefulness, optimized over all instances of the subchannel discrimination problem, is exactly equal to the *robustness of steering*—a natural way of quantifying steering using techniques similar to the ones used in [24], but based on the notion of robustness [44–47]. We argue that the resulting quantification of steering, besides having operational interpretations both in terms of resilience to noise and usefulness, is quantitatively more detailed.

Preliminaries: entanglement and steering.— In the following we will denote by a ˆ (hat) mathematical entities that are "normalized." So, for example, a positive semidefinite operator with unit trace is a (normalized) state $\hat{\rho}$. An *ensemble* $\mathcal{E} = \{\rho_a\}_a$ for a state $\hat{\rho}$ is a collection of *sub*states $\rho_a \leq \hat{\rho}$ such that $\sum_a \rho_a = \hat{\rho}$. Each substate ρ_a can be seen as being proportional to a normalized state $\hat{\rho}_a$, $\rho_a = p_a\hat{\rho}_a$, with $p_a = \mathrm{Tr}(\rho_a)$ being the probability of $\hat{\rho}_a$ in the ensemble. An *assemblage* $\mathcal{A} = \{\mathcal{E}_x\}_x = \{\rho_{a|x}\}_{a,x}$ is a collection of ensembles $\mathcal{E}_x$ for the same state $\hat{\rho}$, one for each x, i.e., $\sum_a \rho_{a|x} = \hat{\rho}$, for all x. For example, $\mathcal{E} = \{\frac{1}{2}|0\rangle\langle 0|, \frac{1}{2}|1\rangle\langle 1|\}$ and $\mathcal{E}' = \{\frac{1}{2}|+\rangle\langle +|, \frac{1}{2}|-\rangle\langle -|\}$, with $|\pm\rangle := (|0\rangle \pm |1\rangle)/\sqrt{2}$, are both ensembles for the maximally mixed state $\mathbb{1}/2$ of a qubit, and taken together they form an assemblage $\mathcal{A} = \{\mathcal{E}, \mathcal{E}'\}$ for $\mathbb{1}/2$.

Along similar lines, a *measurement assemblage* $\mathcal{MA} = \{M_{a|x}\}_{a,x}$ is a collection of positive operators $M_{a|x} \geq 0$ satisfying $\sum_a M_{a|x} = \mathbb{1}$ for each x. Such a collection represents one *positive-operator-valued measure* (or POVM), describing a general quantum measurement, for each x. For a fixed bipartite state $\hat{\rho}_{AB}$, every measurement assemblage on Alice gives rise to an assemblage on Bob:

$$\rho^B_{a|x} = \mathrm{Tr}_A\big(M^A_{a|x}\hat{\rho}_{AB}\big). \tag{1}$$

On the other hand, every assemblage on Bob $\{\sigma_{a|x}\}_{a,x}$ has a quantum realization (1) for some $\hat{\rho}_{AB}$ satisfying $\hat{\rho}_B = \mathrm{Tr}_A(\hat{\rho}_{AB}) = \sum_x \sigma_{a|x} =: \hat{\sigma}_B$ and for some measurement assemblage $\{M_{a|x}\}_{a,x}$ [48].

An assemblage $\mathcal{A} = \{\rho_{a|x}\}_{a,x}$ is *unsteerable* if

$$\rho^{\mathrm{US}}_{a|x} = \sum_\lambda p(\lambda)p(a|x,\lambda)\hat{\sigma}(\lambda) = \sum_\lambda p(a|x,\lambda)\sigma(\lambda), \tag{2}$$

for all a, x, for some probability distribution $p(\lambda)$, conditional probability distributions $p(a|x,\lambda)$, and states $\hat{\sigma}(\lambda)$. Here λ indicates a (hidden) classical random variable, and we introduced also subnormalized states $\sigma(\lambda) = p(\lambda)\hat{\sigma}(\lambda)$. We observe that every conditional probability distribution $p(a|x,\lambda)$ can be written as a convex combination of deterministic conditional probability distributions: $p(a|x,\lambda) = \sum_\nu p(\nu|\lambda)D(a|x,\nu)$, where $D(a|x,\nu) = \delta_{a,f_\nu(x)}$ is a deterministic response function labeled by ν. This means that, by a suitable relabeling,

$$\rho^{\mathrm{US}}_{a|x} = \sum_\lambda D(a|x,\lambda)\sigma(\lambda) \quad \forall a, x, \tag{3}$$

where the summation is over labels of deterministic response functions. We say that an assemblage $\{\rho_{a|x}\}_{a,x}$ is *steerable* if it is not unsteerable.

A *separable* (or *unentangled*) state is one that admits a decomposition

$$\hat{\sigma}^{\mathrm{sep}}_{AB} = \sum_\lambda p(\lambda)\hat{\sigma}_A(\lambda) \otimes \hat{\sigma}_B(\lambda), \tag{4}$$

for $\hat{\sigma}_A(\lambda)$, $\hat{\sigma}_B(\lambda)$ local states, λ a classical label, and $p(\lambda)$ a probability distribution [49]. A state is *entangled* if it is not separable. An unsteerable assemblage can always be obtained via (1) from the separable state $\rho_{AB} = \sum_\lambda p(\lambda)|\lambda\rangle\langle\lambda|_A \otimes \hat{\sigma}(\lambda)_B$, $M_{a|x} = \sum_\mu p(a|x,\mu)|\mu\rangle\langle\mu|$, and $\langle\mu|\lambda\rangle = \delta_{\mu\lambda}$. Most importantly, any separable state can only give raise to unsteerable assemblages. Indeed, for a separable state of the form (4), one has

$$\sigma^{\mathrm{US}}_{a|x} = \mathrm{Tr}_A(M_{a|x}\sigma^{\mathrm{sep}}_{AB}) = \sum_\lambda p(\lambda)p(a|x,\lambda)\sigma_B(\lambda),$$

with $p(a|x,\lambda) = \mathrm{Tr}_A(M_{a|x}\sigma_A(\lambda))$. It follows that entanglement is a necessary condition for steerability, and, in turn, a steerable assemblage is a clear signature of entanglement. Interestingly, not all entangled states lead to steerable assemblages by the action of appropriate local measurement assemblages [4, 5]; we call *steerable states* those that do, and *unsteerable states* those that do not. There exist entangled states that are steerable by one party but not the other (see, e.g., [22]). In this Letter, when we refer to a state being steerable or unsteerable, it is always to be assumed that Alice is the steering party.

Channel and subchannel identification.— A *subchannel* Λ is a linear completely positive map that is trace non-increasing: $\mathrm{Tr}(\Lambda[\rho]) \leq \mathrm{Tr}(\rho)$, for all states ρ. If a subchannel Λ is trace-preserving, $\mathrm{Tr}(\Lambda[\rho]) = \mathrm{Tr}(\rho)$, for all ρ, we use the ˆ notation and say that $\hat{\Lambda}$ is a *channel*. An *instrument* $\mathcal{I} = \{\Lambda_a\}_a$ for a channel $\hat{\Lambda}$ is a collection of subchannels Λ_a such that $\hat{\Lambda} = \sum_a \Lambda_a$ (see Figure 1). Every instrument has (in principle) a physical realization, where the (classical) index a can be considered available to some party [2, 50, 51].

Fix an instrument $\{\Lambda_a\}_a$ for a channel $\hat{\Lambda}$, and consider a measurement $\{Q_b\}_b$ on the output space of $\hat{\Lambda}$. The joint probability of Λ_a and Q_b for input ρ is $p(a,b) := \mathrm{Tr}(Q_b\Lambda_a[\rho]) = p(b|a)p(a)$, where $p(a) = \mathrm{Tr}(\Lambda_a[\rho])$ is the probability of the subchannel Λ_a for the given input ρ and $p(b|a) = p(a,b)/p(a)$ is the conditional probability of the outcome b given that the subchannel Λ_a took place (see Figure 2(a)). The probability of correctly identifying which subchannel was realized is

$$p_{\mathrm{corr}}(\{\Lambda_a\}_a, \{Q_b\}_b, \rho) = \sum_a \mathrm{Tr}(Q_a\Lambda_a[\rho]). \tag{5}$$

The archetypal case of subchannel discrimination is that of *channel* discrimination, where $\Lambda_a = p_a\hat{\Lambda}_a$, with channels $\hat{\Lambda}_a$ and probabilities p_a. The problem often considered is that of telling apart just two channels $\hat{\Lambda}_0$ and $\hat{\Lambda}_1$, each given with probability $p_0 = p_1 = 1/2$. In this case the total (average) channel is simply $\hat{\Lambda} = \frac{1}{2}\hat{\Lambda}_0 + \frac{1}{2}\hat{\Lambda}_1$. The best success probability in identifying subchannels $\{\Lambda_a\}_a$ with an input ρ is defined as $p_{\mathrm{corr}}(\{\Lambda_a\}_a, \rho) := \max_{\{Q_b\}_b} p_{\mathrm{corr}}(\{\Lambda_a\}_a, \{Q_b\}_b, \rho)$. Optimizing also over the input state, one arrives at $p^{\mathrm{NE}}_{\mathrm{corr}}(\{\Lambda_a\}_a) := \max_\rho p_{\mathrm{corr}}(\{\Lambda_a\}_a, \rho)$, where the superscript NE stands for "no entanglement" (see Fig. 2(a)).

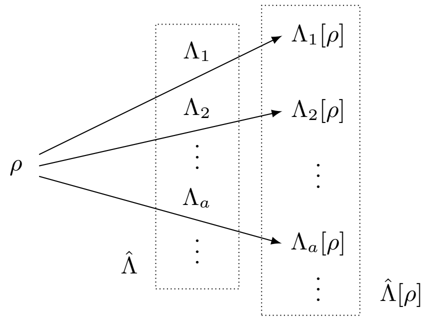

FIG. 1: A decomposition of a channel into subchannels can be seen as a decomposition of a quantum evolution into *branches* of the evolution. If $\{\Lambda_a\}_a$ is an instrument for $\hat{\Lambda}$, then we can imagine that the evolution $\rho \mapsto \hat{\Lambda}[\rho]$ has branches $\rho \mapsto \Lambda_a[\rho]$, where each branch takes place with probability $\mathrm{Tr}(\Lambda_a[\rho])$. The transformation described by the total channel $\hat{\Lambda}$ can be seen as the situation where the "which-branch" information is lost. An example of subchannel discrimination problem is that of distinguishing between the two quantum evolutions $\Lambda_i[\rho] = K_i \rho K_i^\dagger$, $i = 0, 1$, with $K_0 = |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|$ and $K_1 = \sqrt{\gamma}|0\rangle\langle 1|$, corresponding to the so-called amplitude damping channel $\hat{\Lambda} = \Lambda_0 + \Lambda_1$ [2].

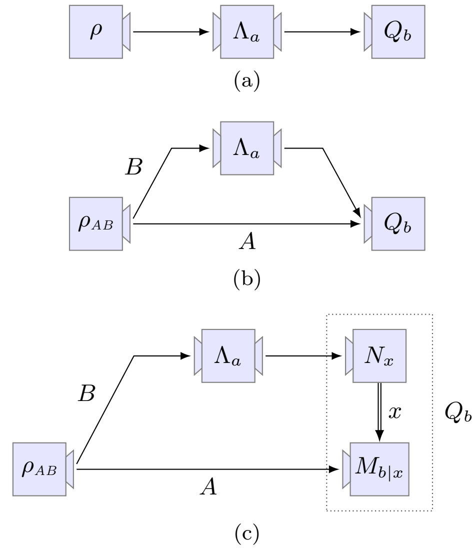

FIG. 2: Different strategies for subchannel discrimination. (a) No entanglement is used: a probe, initially in the state ρ, undergoes the quantum evolution $\hat{\Lambda}$, with branches Λ_a, and is later measured, with an outcome b for the measurement described by the POVM $\{Q_b\}_b$, which is the guess for which branch of the evolution actually took place. (b) The probe B is potentially entangled with an ancilla A; the output probe and the ancilla are jointly measured. (c) The probe is still potentially entangled with an ancilla, but the final measurement $\{Q_b\}_b$ is restricted to local measurements on the output probe and the ancilla, coordinated by one-way classical communication (single lines represent quantum systems, double lines classical information): the outcome x of the measurement $\{N_x\}_x$ performed on the output probe is used to decide which measurement $\{M_{b|x}\}_b$ to perform on the ancilla.

Indeed, one may try to improve the success probability by using an entangled input state ρ_{AB} of an input probe B and an ancilla A. The guess about which subchannel took place is based on a joint measurement of the output probe and the ancilla (see Fig. 2(b)), with success probability $p_{\mathrm{corr}}(\{\Lambda_a^B\}_a, \{Q_b^{AB}\}_b, \rho_{AB})$. In the latter expression we have explicitly indicated that the subchannels act non-trivially only on B, while input state and measurement pertain to AB. One can define the optimal probability of success for a scheme that uses input entanglement and global measurements: $p_{\mathrm{corr}}^{\mathrm{E}}(\{\Lambda_a\}_a) := \max_{\rho_{AB}} \max_{\{Q_b^{AB}\}_b} p_{\mathrm{corr}}(\{\Lambda_a^B\}_a, \{Q_b^{AB}\}_b, \rho_{AB})$. We say that entanglement is useful in discriminating subchannels $\{\Lambda_a\}_a$ if $p_{\mathrm{corr}}^{\mathrm{E}}(\{\Lambda_a\}_a) > p_{\mathrm{corr}}^{\mathrm{NE}}(\{\Lambda_a\}_a)$. It is known that there are instances of subchannel discrimination, already in the simple setting $\{\Lambda_a\}_a = \{\frac{1}{2}\hat{\Lambda}_0, \frac{1}{2}\hat{\Lambda}_0\}$, where $p_{\mathrm{corr}}^{\mathrm{E}} \approx 1 \gg p_{\mathrm{corr}}^{\mathrm{NE}} \approx 0$ (see [43] and references therein).

In [42] it was proven that, for any entangled state ρ_{AB}, there exists a choice $\{\frac{1}{2}\hat{\Lambda}_0, \frac{1}{2}\hat{\Lambda}_1\}$ such that

$$p_{\mathrm{corr}}\left(\left\{\frac{1}{2}\hat{\Lambda}_0, \frac{1}{2}\hat{\Lambda}_1\right\}, \rho_{AB}\right) > p_{\mathrm{corr}}^{\mathrm{NE}}\left(\left\{\frac{1}{2}\hat{\Lambda}_0, \frac{1}{2}\hat{\Lambda}_1\right\}\right),$$

i.e., that every entangled state is useful for the task of (sub)channel discrimination. In this sense, every entangled state, independently of how weakly entangled it is, is a resource. Nonetheless, exploiting such a resource may require arbitrary joint measurements on the output probe and ancilla [43]. From a conceptual perspective, one may want to limit measurements to those that can be performed by local operations and classical communication (LOCC), as this makes the input entangled state the only non-local resource. This limitation can be justified also from a practical perspective: LOCC measurements are arguably easier to implement, and might be the only feasible kind of measurements, especially in a scenario where only weakly entangled states can be produced. We do not know whether every entangled state stays useful for subchannel discrimination when measurements are restricted to be LOCC. In the following, though, we prove that, if the measurements are limited to local operations and forward communication (one-way LOCC), then only steerable states remain useful.

Steerability and subchannel identification by means of restricted measurements.— A Bob-to-Alice one-way LOCC measurement of the form $\mathcal{M}^{B\to A} = \{Q_a^{B\to A}\}_a$ has the structure $Q_a^{B\to A} = \sum_x M_{a|x}^A \otimes N_x^B$, where $\{N_x^B\}_x$ is a measurement on B and $\{M_{a|x}^A\}_{a,x}$ is a measurement assemblage on A. We define $p_{\mathrm{corr}}^{B\to A}(\mathcal{I}, \rho_{AB}) := \max_{\mathcal{M}^{B\to A}} p_{\mathrm{corr}}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \rho_{AB})$ as the optimal probability of success in the discrimination of the instrument $\mathcal{I}^B = \{\Lambda_a^B\}_a$ by means of the input state ρ_{AB} and one-way LOCC measurements from B to A (see Fig. 2(c)). We say that ρ_{AB} is useful in this restricted-measurement scenario if $p_{\mathrm{corr}}^{B\to A}(\mathcal{I}, \rho_{AB}) > p_{\mathrm{corr}}^{\mathrm{NE}}(\mathcal{I})$ for some instrument

$\mathcal{I}$ [59]. Using (1), we find that

$$p_{\rm corr}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \rho_{AB}) = \sum_{a,x} {\rm Tr}_B(\Lambda_a^{\dagger B}[N_x^B]\rho_{a|x}), \quad (6)$$

where $\Lambda_a^\dagger$ denotes the dual map to Λ_a, defined via ${\rm Tr}(X\Lambda_a[Y]) = {\rm Tr}(\Lambda_a^\dagger[X]Y)$, $\forall X, Y$ (assuming Λ_a is completely positive). If the assemblage $\mathcal{A} = \{\rho_{a|x}\}_{a,x}$ appearing in (6) is unsteerable, then we can achieve an equal or better performance with an uncorrelated probe in the best input state $\hat{\sigma}(\lambda)$ among the ones appearing in Eq. (2). Thus, if ρ_{AB} is unsteerable, then it is useless for subchannel discrimination with one-way measurements. This applies also to entangled states that are unsteerable, which are nonetheless useful in channel discrimination with arbitrary measurements [42].

We will now prove that every steerable state *is* useful in subchannel discrimination with one-way-LOCC measurements. To state our result in full detail we need to introduce the *steering robustness of* ρ_{AB},

$$R_{\rm steer}^{A\to B}(\rho_{AB}) := \sup_{\mathcal{MA}} R(\mathcal{A}), \quad (7)$$

where the supremum is over all measurement assemblages $\mathcal{MA} = \{M_{a|x}\}_{a,x}$ on A, $R(\mathcal{A})$ is the *steering robustness of the assemblage* $\mathcal{A}$,

$$R(\mathcal{A}) := \min\left\{ t \geq 0 \,\middle|\, \left\{\frac{\rho_{a|x} + t\,\tau_{a|x}}{1+t}\right\}_{a,x} \text{unsteerable}, \ \{\tau_{a|x}\} \text{ an assemblage}\right\}, \quad (8)$$

and $\mathcal{A}$ is obtained from ρ_{AB} with the measurement assemblage $\mathcal{MA}$ on A (see Eq. (1)). The steering robustness of $\mathcal{A}$ is a measure of the minimal "noise" needed to destroy the steerability of the assemblage $\mathcal{A}$, where such noise is in terms of the mixing with an arbitrary assemblage $\{\tau_{a|x}\}_{a,x}$. With the notation set, we have the following theorem.

Theorem 1. *Every steerable state is useful in one-way subchannel discrimination. More precisely, it holds*

$$\sup_{\mathcal{I}} \frac{p_{corr}^{B\to A}(\mathcal{I}, \rho_{AB})}{p_{corr}^{\rm NE}(\mathcal{I})} = R_{steer}^{A\to B}(\rho_{AB}) + 1, \quad (9)$$

where the supremum is over all instruments $\mathcal{I}$.

Proof. Using the definitions (7) and (8) it is immediate to verify (see Appendix)

$$p_{\rm corr}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \rho_{AB}) \leq (1 + R_{\rm steer}^{A\to B}(\rho_{AB}))p_{\rm corr}^{\rm NE}(\mathcal{I}),$$

for any $\mathcal{M}^{B\to A}$ and any $\mathcal{I}$. We will prove next that the bound can be approximated arbitrarily well. We will do so by constructing appropriate instances of the subchannel discrimination problem. To do this, we will need that the steering robustness $R(\mathcal{A})$ of any assemblage $\mathcal{A} = \{\rho_{a|x}\}_{a,x}$ can be calculated via semidefinite programming (SDP) [52]. In particular, in the Appendix we prove that $R(\mathcal{A}) + 1$ is equal to the optimal value of the SDP optimization problem

$$\text{maximize} \quad \sum_{a,x} {\rm Tr}(F_{a|x}\rho_{a|x}) \quad (10a)$$

$$\text{subject to} \quad \sum_{a,x} D(a|x,\lambda)F_{a|x} \leq \mathbb{1} \quad \forall\lambda \quad (10b)$$

$$F_{a|x} \geq 0 \quad \forall a, x, \quad (10c)$$

where the λ's are labels for the deterministic response functions.

Now, let $\mathcal{MA} = \{M_{a|x}\}_{a,x}$ be a measurement assemblage on A, and $\mathcal{A}$ the resulting assemblage on B. Let $F_{a|x}$ be optimal, i.e., such that $\sum_{a,x} {\rm Tr}(F_{a|x}\rho_{a|x}) = 1 + R(\mathcal{A})$. Define linear maps Λ_a via their duals, as

$$\Lambda_a^\dagger = \Lambda_a^\dagger \circ \Pi_X \qquad \forall a, \quad (11)$$

$$\Lambda_a^\dagger\left[|x\rangle\langle x|\right] = \alpha F_{a|x} \qquad \forall a, x. \quad (12)$$

Here $\circ$ is composition, and Π_X indicates the projector onto an orthonormal basis $\{|x\rangle\}$, $x = 1, \ldots, |X|$, where $|X|$ is the number of settings in the measurement assemblage $\mathcal{MA}$. The constant $\alpha > 0$ will be chosen soon. Because of the conditions (10c), (11), and (12), the $\Lambda_a^\dagger$'s are completely positive linear maps, hence the Λ_a's are too; they act according to $\Lambda_a[\rho] = \alpha \sum_x {\rm Tr}(F_{a|x}\rho)|x\rangle\langle x|$, and can be seen as subchannels as long as $\sum_a \Lambda_a^\dagger[\mathbb{1}] = \sum_{a,x}\Lambda_a^\dagger[|x\rangle\langle x|] = \alpha\sum_{a,x} F_{a|x} \leq \mathbb{1}$, a condition that can be satisfied for $\alpha = \|\sum_{a,x} F_{a|x}\|_\infty^{-1}$, with $\|\cdot\|_\infty$ the operator norm.

We can now introduce N additional subchannels, defined as $\Lambda_a[\rho] = \frac{1}{N}{\rm Tr}((\mathbb{1} - \sum_a \Lambda_a^\dagger[\mathbb{1}])\rho)\hat{\sigma}_a$, for $a = |A| + 1, \ldots, |A| + N$, where $|A|$ indicates the original number of outcomes for POVMs in $\mathcal{MA}$, and $\hat{\sigma}_a$ are arbitrary states in a two-dimensional space orthogonal to ${\rm span}\{|x\rangle \,|\, x = 1, \ldots, |X|\}$. The subchannels Λ_a, $a = 1, \ldots, |A| + N$ do define an instrument $\mathcal{I}$ for the trace-preserving channel $\hat{\Lambda} = \sum_{a=1}^{|A|+N}\Lambda_a$, and one can readily (see Appendix) incorporate the measurement assemblage $\mathcal{MA}$ into a one-way LOCC strategy $\mathcal{M}^{B\to A}$ such that $\alpha(1 + R(\mathcal{A})) \leq p_{\rm corr}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \rho_{AB}) \leq \alpha(1 + R(\mathcal{A})) + \frac{2}{N}$. On the other hand, condition (10b) implies (see Appendix) $\alpha \leq p_{\rm corr}^{\rm NE}(\mathcal{I}) \leq \alpha + \frac{2}{N}$, so $p_{\rm corr}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \rho_{AB})/p_{\rm corr}^{\rm NE}(\mathcal{I}) \geq \frac{1+R(\mathcal{A})}{1+2/(\alpha N)}$. The claim follows since N can be chosen arbitrarily large. □

Conclusions.— We have proven that the steerable states are precisely those states that are useful for the task of subchannel discrimination with feed-forward local measurements. This provides a satisfactorily answer to a question left open by [43] about the characterization of a large class of entangled states that remain useful for (sub)channel discrimination with local measurements. Most importantly, it provides a full operational characterization—and proof of usefulness—of steering in terms of a fundamental task, subchannel discrimination, in a setting—that of restricted measurements—very relevant from the practical point of view. The construction in the proof of Theorem 1 proves that, for any measurement assemblage $\mathcal{MA}$ on A such that the corresponding $\mathcal{A}$ exhibit steering with robustness $R(\mathcal{A}) > 0$, there exist instances of the subchannel discrimination problem with restricted measurements where the use of the steerable state ensures a probability of success approximately $(1 + R(\mathcal{A}))$-fold higher than in the case where no entanglement is used. Thus, the robustnesses $R(\mathcal{A})$ and $R^{A\to B}_{\text{steer}}(\rho_{AB})$ have operational meanings not only in terms of the resilience of steerability versus noise, but in applicative terms. Also, they constitute semi-device-independent lower bounds,

$$R(\mathcal{A}) \leq R^{A\to B}_{\text{steer}}(\rho_{AB}) \leq R_g(\rho_{AB}), \tag{13}$$

on the generalized robustness of entanglement $R_g(\rho_{AB})$ [45, 46], defined as

$$\min\left\{t \geq 0 \,\middle|\, \frac{\rho_{AB} + t\,\tau_{AB}}{1+t} \text{ separable}, \tau \text{ a state}\right\}, \tag{14}$$

which is an entanglement measure with operational interpretations itself [53, 54]. That (13) holds is immediate, given definitions (7) and (8) and the fact that a separable state leads always to unsteerable assemblages. Besides these observations, we believe that the way to quantify steerability that we have introduced is finer-grained than the approach of [24], while preserving the computational efficiency deriving from the use of semidefinite programming. For example, while the so-called *steering weight* of [24] is such that all pure entangled states, however weekly entangled, are deemed maximally steerable, because of (13) we know that weakly entangled pure states have small steering robustness [46]. On the other hand, maximally entangled states ψ_d^+ for large local dimension d do have large steering robustness. Indeed, in the Appendix we prove that, if d is some power of a prime number, then $R^{A\to B}_{\text{steer}}(\psi_d^+) \geq \sqrt{d} - 2$.

Many questions remain open for further investigation: a closed formula for the steerability robustness of pure (maximally entangled) states; whether the result of Theorem 1 can be strengthened to prove that every steerable state is useful for channel—rather than general subchannel—discrimination with restricted measurements; whether general LOCC (rather than one-way LOCC) measurements can restore the usefulness of all entangled states for (sub)channel discrimination.

Acknowledgments.—We acknowledge useful discussions and correspondence with D. Cavalcanti and M. Pusey. We also acknowledge support from NSERC and CIFAR.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009), URL http://link.aps.org/doi/10.1103/RevModPhys.81.865.
[2] M. A. Nielsen and I. L. Chuang (2010).
[3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014), URL http://link.aps.org/doi/10.1103/RevModPhys.86.419.
[4] H. M. Wiseman, S. J. Jones, and A. C. Doherty, Phys. Rev. Lett. **98**, 140402 (2007), URL http://link.aps.org/doi/10.1103/PhysRevLett.98.140402.
[5] S. J. Jones, H. M. Wiseman, and A. C. Doherty, Phys. Rev. A **76**, 052116 (2007), URL http://link.aps.org/doi/10.1103/PhysRevA.76.052116.
[6] M. Born, *The Born-Einstein Letters* (Walker, 1971).
[7] E. Schrödinger, Mathematical Proceedings of the Cambridge Philosophical Society **31**, 555 (1935), ISSN 1469-8064, URL http://journals.cambridge.org/article_S0305004100013554.
[8] A. Einstein, B. Podolsky, and N. Rosen, Physical review **47**, 777 (1935).
[9] E. G. Cavalcanti, M. J. W. Hall, and H. M. Wiseman, Phys. Rev. A **87**, 032306 (2013), URL http://link.aps.org/doi/10.1103/PhysRevA.87.032306.
[10] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002), URL http://link.aps.org/doi/10.1103/RevModPhys.74.145.
[11] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007), URL http://link.aps.org/doi/10.1103/PhysRevLett.98.230501.
[12] C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani, and H. M. Wiseman, Phys. Rev. A **85**, 010301 (2012), URL http://link.aps.org/doi/10.1103/PhysRevA.85.010301.
[13] E. G. Cavalcanti, S. J. Jones, H. M. Wiseman, and M. D. Reid, Phys. Rev. A **80**, 032112 (2009), URL http://link.aps.org/doi/10.1103/PhysRevA.80.032112.
[14] D. J. Saunders, S. J. Jones, H. M. Wiseman, and G. J. Pryde, Nat Phys **6**, 845 (2010), URL http://dx.doi.org/10.1038/nphys1766.
[15] Z. Y. Ou, S. F. Pereira, H. J. Kimble, and K. C. Peng, Phys. Rev. Lett. **68**, 3663 (1992), URL http://link.aps.org/doi/10.1103/PhysRevLett.68.3663.
[16] W. P. Bowen, R. Schnabel, P. K. Lam, and T. C. Ralph, Phys. Rev. Lett. **90**, 043601 (2003), URL http://link.aps.org/doi/10.1103/PhysRevLett.90.043601.
[17] D. H. Smith, G. Gillett, M. P. de Almeida, C. Branciard, A. Fedrizzi, T. J. Weinhold, A. Lita, B. Calkins, T. Gerrits, H. M. Wiseman, et al., Nature communications **3**, 625 (2012).
[18] A. J. Bennet, D. A. Evans, D. J. Saunders, C. Branciard, E. G. Cavalcanti, H. M. Wiseman, and G. J. Pryde, Phys. Rev. X **2**, 031003 (2012), URL http://link.aps.org/doi/10.1103/PhysRevX.2.031003.
[19] V. Händchen, T. Eberle, S. Steinlechner, A. Samblowski, T. Franz, R. F. Werner, and R. Schnabel, Nature Photonics **6**, 596 (2012).
[20] S. Steinlechner, J. Bauchrowitz, T. Eberle, and R. Schnabel, Phys. Rev. A **87**, 022104 (2013), URL http://link.aps.org/doi/10.1103/PhysRevA.87.022104.

[21] B. Wittmann, S. Ramelow, F. Steinlechner, N. K. Langford, N. Brunner, H. M. Wiseman, R. Ursin, and A. Zeilinger, New Journal of Physics **14**, 053030 (2012).

[22] J. Bowles, T. Vértesi, M. T. Quintino, and N. Brunner, Physical Review Letters **112**, 200402 (2014).

[23] M. F. Pusey, Phys. Rev. A **88**, 032313 (2013), URL `http://link.aps.org/doi/10.1103/PhysRevA.88.032313`.

[24] P. Skrzypczyk, M. Navascués, and D. Cavalcanti, Phys. Rev. Lett. **112**, 180404 (2014), URL `http://link.aps.org/doi/10.1103/PhysRevLett.112.180404`.

[25] S. Jevtic, M. F. Pusey, D. Jennings, and T. Rudolph, arXiv preprint arXiv:1303.4724 (2013).

[26] A. Milne, S. Jevtic, D. Jennings, H. Wiseman, and T. Rudolph, arXiv preprint arXiv:1403.0418 (2014).

[27] T. Moroder, O. Gittsovich, M. Huber, and O. Gühne, arXiv preprint arXiv:1405.0262 (2014).

[28] J. Schneeloch, C. J. Broadbent, S. P. Walborn, E. G. Cavalcanti, and J. C. Howell, Physical Review A **87**, 062103 (2013).

[29] J. Schneeloch, P. B. Dixon, G. A. Howland, C. J. Broadbent, and J. C. Howell, Physical review letters **110**, 130407 (2013).

[30] J. Schneeloch, C. J. Broadbent, and J. C. Howell, Physics Letters A (2014).

[31] A. Kitaev, Russ. Math. Surv. **52**, 1191 (1997).

[32] V. Paulsen, *Completely Bounded Maps and Operator Algebras*, Cambridge Studies in Advanced Mathematics (Cambridge University Press, 2002).

[33] A. Childs, J. Preskill, and J. Renes, J. Mod. Opt. **47**, 155 (2000).

[34] G. M. D'Ariano, P. LoPresti, and M. G. A. Paris, Phys. Rev. Lett. **87**, 270404 (2001).

[35] A. Acin, Phys. Rev. Lett. **87**, 177901 (2001).

[36] V. Giovannetti, S. Lloyd, and L. Maccone, Science **306**, 1330 (2004).

[37] A. Gilchrist, N. K. Langford, and M. A. Nielsen, Phys. Rev. A **71**, 062310 (2005).

[38] B. Rosgen and J. Watrous, in *Proc. 20th Ann. Conf. Comp. Compl.* (2005), pp. 344–354.

[39] M. F. Sacchi, Phys. Rev. A **71**, 062340 (2005).

[40] M. F. Sacchi, Phys. Rev. A **72**, 014305 (2005).

[41] S. Lloyd, Science **321**, 1463 (2008).

[42] M. Piani and J. Watrous, Phys. Rev. Lett. **102**, 250501 (pages 4) (2009).

[43] W. Matthews, M. Piani, and J. Watrous, Phys. Rev. A **82**, 032302 (2010), URL `http://link.aps.org/doi/10.1103/PhysRevA.82.032302`.

[44] G. Vidal and R. Tarrach, Phys.Rev. A **59**, 141 (1999).

[45] M. Steiner, Phys. Rev. A **67**, 054305 (2003).

[46] A. Harrow and M. Nielsen, Phys. Rev. A **68**, 012308 (2003).

[47] J. Geller and M. Piani, arXiv preprint arXiv:1401.8197 (2014).

[48] L. P. Hughston, R. Jozsa, and W. K. Wootters, Physics Letters A **183**, 14 (1993), ISSN 0375-9601, URL `http://www.sciencedirect.com/science/article/pii/0375960193908809`.

[49] R. F. Werner, Phys. Rev. A **40**, 4277 (1989), URL `http://link.aps.org/doi/10.1103/PhysRevA.40.4277`.

[50] E. B. Davies and J. T. Lewis, Communications in Mathematical Physics **17**, 239 (1970).

[51] M. Horodecki, Quantum information and computation **1**, 3 (2001).

[52] S. P. Boyd and L. Vandenberghe, *Convex optimization* (Cambridge university press, 2004).

[53] F. G. S. L. Brandão and N. Datta, IEEE Transactions on Information Theory **57**, 1754 (2011).

[54] F. G. S. L. Brandão, Phys. Rev. A **76**, 030301(R) (2007).

[55] G. Vidal and R. F. Werner, Phys. Rev. A **65**, 032314 (2002), URL `http://link.aps.org/doi/10.1103/PhysRevA.65.032314`.

[56] Z. Jiang, M. Piani, and C. M. Caves, Quantum information processing **12**, 1999 (2013).

[57] I. Bengtsson, arXiv preprint quant-ph/0610216 (2006).

[58] R. Jozsa and J. Schlienz, Phys. Rev. A **62**, 012301 (2000).

[59] Notice that no bipartite state ρ_{AB} is useful in one-way subchannel identification when the communication goes from the ancilla to the output probe. This is because the initial measurement of the ancilla simply creates an ensemble of input substates for the channel, and we might as well choose the best input to begin with. So, the only one-way communication that may have a non-trivial effect is that from the output probe to the ancilla.

Robustness as semidefinite program

Inspired by the work of Pusey [23] and Skrzypczyk et al. [24], we are going to prove that calculating the steering robustness $R(\mathcal{A})$ of an assemblage $\mathcal{A}=\{\rho_{a|x}\}_{a,x}$ falls under the umbrella of semidefinite programming (SDP) [52].

By definition, see Eq. (8), $R(\mathcal{A})$ is the minimum positive t such that

$$\rho_{a|x}=(1+t)\sigma^{\mathrm{US}}_{a|x}-t\tau_{a|x},\quad \forall a,x,$$

with $\{\sigma^{\mathrm{US}}_{a|x}\}_{a,x}$ an unsteerable assemblage and $\{\tau_{a|x}\}_{a,x}$ an arbitrary assemblage. Notice that, since $\{\rho_{a|x}\}_{a,x}$ and $\{\sigma^{\mathrm{US}}_{a|x}\}_{a,x}$ are assemblages, $\tau_{a|x}=\big((1+t)\sigma^{\mathrm{US}}_{a|x}-\rho_{a|x}\big)/t$ is automatically an assemblage as long as

$$(1+t)\sigma^{\mathrm{US}}_{a|x}\geq\rho_{a|x},,\quad\forall a,x, \tag{15}$$

Since $\{\sigma^{\mathrm{US}}_{a|x}\}_{a,x}$ is unsteerable, see Eq. (3), we can rewrite Eq. (15) as the condition

$$(1+t)\sum_\lambda D(a|x,\lambda)\sigma_\lambda\geq\rho_{a|x},\quad\forall a,x,$$

where the σ_λ's are subnormalized states, and the sum is over all the deterministic strategies to output a given x. If we consider that the factor $(1+t)$ can be absorbed into the σ_λ's (so that they are generally unnormalized, rather subnormalized), we realize that $R(\mathcal{A})+1$ can be characterized as the solution to

$$\begin{aligned}\text{minimize}\quad & \sum_\lambda \mathrm{Tr}(\sigma_\lambda)\\ \text{subject to}\quad & \sum_\lambda D(a|x,\lambda)\sigma_\lambda\geq\rho_{a|x}\quad\forall a,x\\ & \sigma_\lambda\geq 0\quad\forall\lambda\end{aligned} \tag{16}$$

This is an example of SDP optimization problem [52]. For our purposes, the *primal problem* of an SDP is an optimization problem cast as

$$\begin{aligned} \text{minimize} \quad & \langle C, X\rangle \\ \text{subject to} \quad & \Phi[X] \geq B \\ & X \geq 0, \end{aligned}$$

where:

- $\langle C, X\rangle$ is the objective function;
- B and C are given Hermitian matrices;
- X is the matrix variable on which to optimize;
- $\langle X, Y\rangle := \mathrm{Tr}(X^\dagger Y)$ is the Hilbert-Schmidt inner product;
- Φ is a given Hermiticity-preserving linear map.

The *dual problem* provides a lower bound to the objective function of the primal problem. The dual problem is given by

$$\begin{aligned} \text{maximize} \quad & \langle B, Y\rangle \\ \text{subject to} \quad & \Phi^\dagger[Y] \leq C \\ & Y \geq 0, \end{aligned}$$

where $\Phi^\dagger$ is the dual of Φ with respect to the Hilbert-Schmidt inner product, and Y is another matrix variable.

One says that *strong duality* holds when the optimal values of the primal and dual problems coincide. Strong duality holds in many cases, and in particular under the Slater conditions that (i) the primal and dual problems are both feasible, and moreover the primal problem is *strictly feasible*, meaning that there is a positive definite $X > 0$ such that $\Phi[X] > B$, or (ii) the primal and dual problems are both feasible, and moreover the dual problem is strictly feasible, meaning that there is a $Y > 0$ such that $\Phi^\dagger[Y] < C$. In case (i), not only do the primal and dual values coincide, but there must exist Y_{opt} that achieves the optimal value for the dual problem; and similarly, in the case (ii), there must exist X_{opt} that achieves the optimal value in the primal problem.

In our case

$$C = \mathbb{1}, \quad B = \mathrm{diag}(\rho_{a|x})_{a,x},$$
$$\Phi[X] = \mathrm{diag}\left(\sum_\lambda D(a|x,\lambda) X_\lambda\right)_{a,x}$$

where $\mathrm{diag}(\cdot)_{a,x}$ indicates a block-diagonal matrix whose diagonal blocks are labeled by a, x, and the X_λ's are the diagonal blocks of X, labeled by λ. Thus, we have $\Phi^\dagger[Y] = \mathrm{diag}\left(\sum_{a,x} D(a|x,\lambda) Y_{a|x}\right)_\lambda$, and the dual of the primal problem (16) reads

$$\text{maximize} \quad \sum_{a,x} \mathrm{Tr}(F_{a|x}\rho_{a|x}) \tag{17a}$$
$$\text{subject to} \quad \sum_{a,x} D(a|x,\lambda) F_{a|x} \leq \mathbb{1} \quad \forall \lambda \tag{17b}$$
$$F_{a|x} \geq 0 \quad \forall a, x, \tag{17c}$$

It is easy to verify that both Slater conditions hold in our case. For instance, one can take $\sigma_\lambda = 2\mathbb{1}$ for all λ, and $F_{a|x} = \frac{\mathbb{1}}{|X|+1}$ for all a, x, with $|X|$ being the number of possible values for x. Thus, there exist $F_{a|x} = F^{\mathrm{opt}}_{a|x}$ satisfying the constraints of Eq. (17) and such that $\sum_{a,x} \mathrm{Tr}(F_{a|x}\rho_{a|x}) = 1 + R(\mathcal{A})$.

We remark that the optimal $F_{a|x}$ can always be chosen to saturate (17b). That is, there is a deterministic strategy $D(a|x,\lambda)$ and a normalized pure state $|\phi\rangle$ such that

$$\sum_{a,x} D(a|x,\lambda)\langle\phi|F_{a|x}|\phi\rangle = \langle\phi|\mathbb{1}|\phi\rangle = 1 \tag{18}$$

This is because otherwise it is always possible to increase (in operator sense) some $F_{a|x}$'s, still maintaining the optimal value for the objective function (which is operator monotone in the $F_{a|x}$'s).

Details of the proof of Theorem 1

The claimed upper bound,

$$p_{\mathrm{corr}}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \rho_{AB}) \leq (1 + R^{A\to B}_{\mathrm{steer}}(\rho_{AB}))p^{\mathrm{NE}}_{\mathrm{corr}}(\mathcal{I}),$$

can be proved using (6) and definitions (7) and (8):

$$\begin{aligned} &p_{\mathrm{corr}}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \rho_{AB}) \\ &\quad = \sum_{a,x} \mathrm{Tr}_B(\Lambda^{\dagger B}_a[N^B_x]\rho_{a|x}) \\ &\quad \leq (1 + R(\mathcal{A})) \sum_{a,x} \mathrm{Tr}_B(\Lambda^{\dagger B}_a[N^B_x]\sigma^{\mathrm{US}}_{a|x}) \\ &\qquad - R(\mathcal{A}) \sum_{a,x} \mathrm{Tr}_B(\Lambda^{\dagger B}_a[N^B_x]\tau^{\mathrm{US}}_{a|x}) \\ &\quad \leq (1 + R(\mathcal{A}))p^{\mathrm{NE}}_{\mathrm{corr}}(\mathcal{I}) \\ &\quad \leq (1 + R^{A\to B}_{\mathrm{steer}}(\rho_{AB}))p^{\mathrm{NE}}_{\mathrm{corr}}(\mathcal{I}). \end{aligned}$$

On the other hand, suppose that $\mathcal{MA} = \{M_{a|x}\}_{a,x}$, where $a = 1, \ldots, |A|$ and $x = 1, \ldots, |X|$, is a measurement assemblage on A such that the corresponding assemblage $\mathcal{A} = \{\rho_{a|x} = \mathrm{Tr}_A(M^A_{a|x}\rho_{AB})\}_{a,x}$ is steerable. Let $F_{a|x} \geq 0$ be the operators optimal for (17), such that $\sum_{a,x} \mathrm{Tr}(F_{a|x}\rho_{a|x}) = 1 + R(\mathcal{A})$. In the proof of Theorem 1 of the main text we defined subchannels Λ_a that act as

$$\Lambda_a[\rho] = \begin{cases} \alpha \sum_{x=1}^{|X|} \mathrm{Tr}(\rho F_{a|x})|x\rangle\langle x| & 1 \le a \le |A| \\ \frac{1}{N} \mathrm{Tr}((\mathbb{1} - \sum_{a=1}^{|A|} \Lambda_a^\dagger[\mathbb{1}])\rho)\hat{\sigma}_a & |A|+1 \le a \le |A|+N, \end{cases} \tag{19}$$

where $\alpha = \| \sum_{a,x} F_{a|x}\|_\infty^{-1} > 0$, and the $\hat{\sigma}_a$, $a = |A|+1,\dots,|A|+N$, are arbitrary (normalized) states in a two-dimensional subspace orthogonal to $\mathrm{span}\{|x\rangle \,|\, x = 1,\dots,|X|\}$. It is immediate to check that $\mathrm{Tr}\left(\sum_{a=1}^{|A|+N} \Lambda_a[\rho]\right) = \mathrm{Tr}(\rho)$ (by construction), so $\mathcal{I} = \{\Lambda_a\}_{a=1,\dots,|A|+N}$ is an instrument for the channel $\sum_{a=1}^{|A|+N} \Lambda_a$.

Let σ_{AB} be an arbitrary bipartite state on AB, and let $\mathcal{MA}^{B\to A} = \{Q_a\}_a^{B\to A}$ be an arbitrary one-way measurement from B to A, i.e., $Q_a^{B\to A} = \sum_y M'^A_{a|y} \otimes N'^B_y$, to guess which subchannel was actually realized. Notice that y in the latter expression potentially varies in an arbitrary range, different from the range $\{1,\dots,|X|\}$ for the parameter x of the fixed measurement assemblage $\mathcal{MA}$. Nonetheless we observe that $\Lambda_a = \Pi'_X \circ \Lambda_a$ for $a = 1,\dots,|A|+N$, where $\circ$ is composition, and

$$\Pi'_X[\tau] = \sum_{x=1}^{|X|} |x\rangle\langle x|\tau|x\rangle\langle x| + \Pi^\perp \tau \Pi^\perp,$$

with $\Pi^\perp$ the projector onto the two-dimensional space orthogonal to $\mathrm{span}\{|x\rangle \,|\, x = 1,\dots,|X|\}$ that supports the arbitrary qubits states $\hat{\sigma}_a$, $a = |A|+1,\dots,|A|+N$. Also,

$$\Lambda_a^B[\sigma_{AB}] = \frac{1}{N}\left(\sigma_A - \sum_{a'=1}^{|A|} \mathrm{Tr}_B(\Lambda_{a'}^B[\sigma_{AB}])\right) \otimes \hat{\sigma}_a^B,$$

for $a = |A|+1,\dots,|A|+N$. This implies that, for whatever input σ_{AB}, the optimal $Q_a^{B\to A}$ can be chosen to have the form

$$Q_a^{B\to A} = \begin{cases} \sum_{x=1}^{|X|} M'^A_{a|x} \otimes |x\rangle\langle x|^B & 1 \le a \le |A| \\ \mathbb{1}^A \otimes N_a^B, & |A|+1 \le a \le |A|+N, \end{cases} \tag{20}$$

with $\Pi^\perp N_a \Pi^\perp = N_a$, for $|A|+1 \le a \le |A|+N$, a POVM on the orthogonal qubit space. Omitting a detailed and straightforward proof of this, we instead provide the following intuition: For the subchannels (19), the best local measurement on the output probe is one that first of all discriminates between the space $\mathrm{span}\{|x\rangle \,|\, x = 1,\dots,|X|\}$ and the orthogonal qubit space. If the probe is found in the space $\mathrm{span}\{|x\rangle \,|\, x = 1,\dots,|X|\}$, the probe is then measured in the basis $\{|x\rangle \,|\, x = 1,\dots,|X|\}$ and the result if forwarded to decide which measurement to perform on the ancilla: this is optimal because, in this subspace, the output probe is already dephased in the basis $\{|x\rangle \,|\, x = 1,\dots,|X|\}$. If the probe is instead found in the orthogonal qubit space, there is no information to be gained from the ancilla, since, for the state of the probe to have support in the orthogonal qubit space, the probe must have been discarded and prepared in one of the random qubits states $\hat{\sigma}_a$. So, in this case, the ancilla is necessarily decorrelated and its state independent of the specific Λ_a, $a = |A|+1,\dots,|A|+N$, that has been realized; thus the optimal guess about said Λ_a can be made as soon as the output probe is measured.

Then, for an optimal $\mathcal{M}^{B\to A} = \{Q_a^{B\to A}\}_a$ of the form (20), we find in general

$$\begin{aligned} &p_{\mathrm{corr}}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \sigma_{AB}) \\ &= \sum_{a=1}^{|A|+N} \mathrm{Tr}(Q_a^{B\to A}\Lambda_a^B[\sigma_{AB}]) \\ &= \sum_{a=1}^{|A|} \mathrm{Tr}(Q_a^{B\to A}\Lambda_a^B[\sigma_{AB}]) + \sum_{a=|A|+1}^{|A|+N} \mathrm{Tr}(Q_a^{B\to A}\Lambda_a^B[\sigma_{AB}]) \\ &= \sum_{a=1}^{|A|}\sum_{x=1}^{|X|} \mathrm{Tr}(M'^A_{a|x} \otimes |x\rangle\langle x|_B \Lambda_a^B[\sigma_{AB}]) \\ &\quad + \left(1 - \sum_{a=1}^{|A|} \mathrm{Tr}(\Lambda_a^B[\sigma_{AB}])\right)\frac{1}{N}\sum_{a=|A|+1}^{|A|+N} \mathrm{Tr}(N_a\hat{\sigma}_a) \\ &= \sum_{a=1}^{|A|}\sum_{x=1}^{|X|} \mathrm{Tr}(\Lambda_a^\dagger[|x\rangle\langle x|]\sigma_{a|x}) \\ &\quad + \left(1 - \sum_{a=1}^{|A|} \mathrm{Tr}(\Lambda_a^B[\sigma_{AB}])\right)\frac{1}{N}\sum_{a=|A|+1}^{|A|+N} \mathrm{Tr}(N_a\hat{\sigma}_a), \end{aligned}$$

with $\sigma_{a|x} = \mathrm{Tr}_A(M'_{a|x}\sigma_{AB})$. By construction it holds that $\Lambda_a^\dagger[|x\rangle\langle x|] = \alpha F_{a|x}$ for $1 \le a \le |A|$ and $1 \le x \le |X|$ (see Eq. (12)), therefore

$$\begin{aligned} &p_{\mathrm{corr}}(\mathcal{I}^B, \mathcal{M}^{B\to A}, \rho_{AB}) \\ &= \alpha \sum_{a=1}^{|A|}\sum_{x=1}^{|X|} \mathrm{Tr}(F_{a|x}\sigma_{a|x}) \\ &\quad + \left(1 - \sum_{a=1}^{|A|} \mathrm{Tr}(\Lambda_a^B[\sigma_{AB}])\right)\frac{1}{N}\sum_{a=|A|+1}^{|A|+N} \mathrm{Tr}(N_a\hat{\sigma}_a) \\ &\le \alpha \sum_{a=1}^{|A|}\sum_{x=1}^{|X|} \mathrm{Tr}(F_{a|x}\sigma_{a|x}) + \frac{2}{N}. \end{aligned} \tag{21}$$

In the last line we used

$$\left(1 - \sum_{a=1}^{|A|} \mathrm{Tr}(\Lambda_a^B[\sigma_{AB}])\right) \le 1$$

and

$$\frac{1}{N}\sum_{a=|A|+1}^{|A|+N}\mathrm{Tr}(N_a\hat{\sigma}_a)\leq\frac{1}{N}\mathrm{Tr}\left(\sum_{a=|A|+1}^{|A|+N}N_a\right)$$
$$\leq\frac{1}{N}\mathrm{Tr}(\Pi^{\perp})=\frac{2}{N}. \quad (22)$$

It is clear that if $\sigma_{AB}=\rho_{AB}$ and $M'_{a|x}=M_{a|x}$ in (20), so that $\sigma_{a|x}=\rho_{a|x}$, then we have

$$1+R(\mathcal{A})\leq p_{\mathrm{corr}}(\mathcal{I}^B,\mathcal{M}^{B\to A},\rho_{AB})\leq 1+R(\mathcal{A})+\frac{2}{N}.$$

It remains to prove that

$$\alpha\leq p_{\mathrm{corr}}^{\mathrm{NE}}(\mathcal{I})\leq\alpha+\frac{2}{N}. \quad (23)$$

This is readily verified by considering that (17b) can be saturated, as argued at the end of the previous section (see (18)), for an optimal solution of the SDP problem. So we have that for some deterministic $D(a|x,\lambda)$ and some uncorrelated input state $|\phi\rangle$ to the channel,

$$\begin{aligned}1&=\sum_{a=1}^{|A|}\sum_{x=1}^{|X|}D(a|x,\lambda)\langle\phi|F_{a|x}|\phi\rangle\\&=\frac{1}{\alpha}\sum_{a=1}^{|A|}\sum_{x=1}^{|X|}D(a|x,\lambda)\langle\phi|\Lambda_a^{\dagger}[|x\rangle\langle x|]|\phi\rangle\\&=\frac{1}{\alpha}\sum_{a=1}^{|A|}\mathrm{Tr}\left(\left(\sum_{x:D(a|x,\lambda)=1}|x\rangle\langle x|\right)\Lambda_a[|\phi\rangle\langle\phi|]\right)\\&=\frac{1}{\alpha}\sum_{a=1}^{|A|}\mathrm{Tr}\left(M''_a\Lambda_a[|\phi\rangle\langle\phi|]\right),\end{aligned}$$

having defined $M''_a:=\sum_{x:D(a|x,\lambda)=1}|x\rangle\langle x|$. Considering also the subchannels Λ_a, $a=|A|+1,\ldots,|A|+N$, and bounding their contribution to the probability of success as in (21), we arrive at (23).

On the scaling of the steerability of maximally entangled states

We have argued that $R_{\mathrm{steer}}^{A\to B}(\rho_{AB})\leq R_g(\rho_{AB})$, where $R_g(\rho_{AB})$ is the generalized entanglement robustness (14). Indeed, let τ_{AB} be optimal for the generalized entanglement robustness, i.e., suppose

$$\sigma_{AB}=\frac{\rho_{AB}+R_g(\rho_{AB})\tau_{AB}}{1+R_g(\rho_{AB})}$$

is separable. Then $\sigma_{a|x}=\mathrm{Tr}_A(M_{a|x}\sigma_{AB})$ is unsteerable for any measurement assemblage $\{M_{a|x}\}_{a,x}$, proving that $R_g(\rho_{AB})$ is an upper bound to $R_{\mathrm{steer}}^{A\to B}(\rho_{AB})$ (see Eq. (7)). This means that, if a state is weakly entangled with respect to R_g, it is also weakly steerable with respect to $R_{\mathrm{steer}}^{A\to B}$. In [46] it was proven that, for any bipartite pure state

$$|\psi\rangle_{AB}=\sum_i\sqrt{p_i}|i\rangle_A|i\rangle_B,$$

here in its Schmidt decomposition, the generalized entanglement robustness is equal to

$$R_g(|\psi\rangle\langle\psi|_{AB})=\left(\sum_i\sqrt{p_i}\right)^2-1=2\mathcal{N}(|\psi\rangle\langle\psi|_{AB}),$$

where $\mathcal{N}$ is the negativity of entanglement [55]. In particular, then, for a maximally entangled state in dimension $d\times d$, $|\psi_d^+\rangle_{AB}=\frac{1}{\sqrt{d}}\sum_{i=1}^d|i\rangle_A|i\rangle_B$, one has

$$R_{\mathrm{steer}}^{A\to B}(\psi_{d,AB}^+)\leq R_g(\psi_{d,AB}^+)=d-1,$$

having used the notation $\psi_{d,AB}^+=|\psi_d^+\rangle\langle\psi_d^+|_{AB}$.

We conclude by providing a lower bound on $R_{\mathrm{steer}}^{A\to B}(\psi_{d,AB}^+)$ for d a power of a prime number. We will use techniques similar to the ones used in the examples of [56].

Fix d to be the power of a prime number. Then we know that there there are $d+1$ mutually unbiased bases, i.e., $d+1$ orthonormal sets $\{|\psi_{a|x}\rangle\}_{a=1,\ldots,d}$, one for each $x=1,\ldots,d+1$, such that [57]

$$|\langle\psi_{a|x}|\psi_{b|y}\rangle|=\begin{cases}\delta_{a,b} & x=y\\ \frac{1}{\sqrt{d}} & x\neq y\end{cases}$$

We will consider a measurement assemblage $\{M_{a|x}=|\psi_{a|x}\rangle\langle\psi_{a|x}|\}_{a,x}$. Suppose $\rho_{AB}=\psi_{d,AB}^+$. We have

$$\rho_{a|x}^B=\mathrm{Tr}_A(M_{a|x}^A\psi_{d,AB}^+)=\frac{1}{d}|\psi_{a|x}^*\rangle\langle\psi_{a|x}^*|$$

Here $|\psi_{a|x}^*\rangle$ indicates orthonormal vectors whose coefficients in the local basis $\{|i\rangle_B\}$ are the complex conjugate of the coefficients of $|\psi_{a|x}\rangle$ in the local basis $\{|i\rangle_A\}$. Thus, the bases $\{|\psi_{a|x}^*\rangle\}_{a=1,\ldots,d}$ are still mutually unbiased.

We want to lower bound the steering robustness of $\{\rho_{a|x}^B\}_{a,x}$, which in turn will give us a lower bound on $R_{\mathrm{steer}}^{A\to B}(\psi_{d,AB}^+)$. To do this, we use a specific choice for the $F_{a|x}$'s in (17). We choose $F_{a|x}=\beta|\psi_{a|x}^*\rangle\langle\psi_{a|x}^*|$, where $\beta>0$ will be fixed to satisfy (17b) (condition (17c) is satisfied for any $\beta\geq 0$), i.e.,

$$\left\|\sum_{a,x}D(a|x,\lambda)F_{a|x}\right\|_\infty\leq 1$$

for all deterministic $D(a|x,\lambda)$. With our choice of $F_{a|x}$, this can be achieved by taking

$$\beta\leq\left(\max_\lambda\left\|\sum_x|\psi_{f_\lambda(x)|x}^*\rangle\langle\psi_{f_\lambda(x)|x}^*|\right\|_\infty\right)^{-1} \quad (24)$$

where the maximum is over all functions $f_\lambda : \{1, \ldots, d+1\} \to \{1, \ldots, d\}$, labeled by λ. To estimate the right hand side of (24), we will use the fact [58] that, for

$$|\gamma\rangle_{CD} = \sum_{x=1}^{d+1} |\psi^*_{f_\lambda(x)|x}\rangle_C |x\rangle_D,$$

where $\{|x\rangle\}_{x=1,\ldots,d+1}$ is an orthonormal basis, the spectrum of

$$\mathrm{Tr}_D(|\gamma\rangle\langle\gamma|_{CD}) = \sum_x |\psi^*_{f_\lambda(x)|x}\rangle\langle\psi^*_{f_\lambda(x)|x}|.$$

is the same as the spectrum of

$$\begin{aligned}\mathrm{Tr}_C(|\gamma\rangle\langle\gamma|_{CD}) &= \sum_{x,y} \langle\psi^*_{f_\lambda(x)|x}|\psi^*_{f_\lambda(y)|y}\rangle |y\rangle\langle x| \\ &= \sum_x |x\rangle\langle x| + \frac{1}{\sqrt{d}} \sum_{x\neq y} e^{i\phi_{x,y}} |y\rangle\langle x| \\ &= \left(1 - \frac{1}{\sqrt{d}}\right) \mathbb{1} + \frac{1}{\sqrt{d}} \sum_{x,y} e^{i\phi_{x,y}} |y\rangle\langle x|\end{aligned}$$

where $\phi_{x,y}$ are real numbers representing phases. Thus, we have

$$\begin{aligned}&\left\| \sum_x |\psi^*_{f_\lambda(x)|x}\rangle\langle\psi^*_{f_\lambda(x)|x}| \right\|_\infty \\ &= \left\| \sum_{x,y} \langle\psi^*_{f_\lambda(x)|x}|\psi^*_{f_\lambda(y)|y}\rangle |y\rangle\langle x| \right\|_\infty \\ &\leq \left(1 - \frac{1}{\sqrt{d}}\right) + \frac{1}{\sqrt{d}} \left\| \sum_{x,y} e^{i\phi_{x,y}} |y\rangle\langle x| \right\|_\infty \\ &\leq \left(1 - \frac{1}{\sqrt{d}}\right) + \frac{1}{\sqrt{d}} \left\| \sum_{x,y} e^{i\phi_{x,y}} |y\rangle\langle x| \right\|_2 \\ &= \left(1 - \frac{1}{\sqrt{d}}\right) + \frac{1}{\sqrt{d}} (d+1) \\ &= 1 + \sqrt{d}.\end{aligned}$$

Since this estimate is independent of λ, we can take $\beta = 1/(\sqrt{d}+1)$. Hence, we conclude that, for d the power of a prime number,

$$\begin{aligned}&R^{A\to B}_{\mathrm{steer}}(\psi^+_{d,AB}) \\ &\geq R\left(\left\{\frac{1}{d}|\psi^*_{a|x}\rangle\langle\psi^*_{a|x}|\right\}\right) \\ &\geq \sum_{a,x} \mathrm{Tr}\left(\left(\frac{1}{d}|\psi^*_{a|x}\rangle\langle\psi^*_{a|x}|\right)\left(\frac{1}{\sqrt{d}+1}|\psi^*_{a|x}\rangle\langle\psi^*_{a|x}|\right)\right) - 1 \\ &= \frac{1}{d(\sqrt{d}+1)}(d(d+1)) - 1 \\ &= \sqrt{d}\frac{\sqrt{d}-1}{\sqrt{d}+1} \\ &\geq \sqrt{d} - 2.\end{aligned} \tag{25}$$