Quantum Steering Beyond Instrumental Causal Networks

4 authors, including:

Ranieri Vieira Nery
Federal University of Rio de Janeiro
7 PUBLICATIONS   **74** CITATIONS

Quantum steering beyond instrumental causal networks

R. V. Nery,[1] M. M. Taddei,[1] R. Chaves,[2] and L. Aolita[1,3]

[1]*Instituto de Física, Universidade Federal do Rio de Janeiro,*
P. O. Box 68528, Rio de Janeiro, RJ 21941-972, Brazil
[2]*International Institute of Physics, Federal University of Rio Grande do Norte, 59078-970, P. O. Box 1613, Natal, Brazil*
[3]*ICTP South American Institute for Fundamental Research Instituto de Física Teórica,*
UNESP-Universidade Estadual Paulista R. Dr. Bento T. Ferraz 271, Bl. II, São Paulo 01140-070, SP, Brazil

We theoretically predict, and experimentally verify with entangled photons, that outcome communication is not enough for hidden-state models to reproduce quantum steering. Hidden-state models with outcome communication correspond, in turn, to the well-known *instrumental processes* of causal inference but in the 1-sided device-independent (1S DI) scenario of one black-box measurement device and one well-characterised quantum apparatus. We introduce *1S-DI instrumental inequalities* to test against these models, with the appealing feature of detecting entanglement even when communication of the black box's measurement outcome is allowed. We find that, remarkably, these inequalities can also be violated solely with steering, i.e. without outcome communication. In fact, an efficiently-computable formal quantifier – the *robustness of non-instrumentality* – naturally arises; and we prove that steering alone is enough to maximize it. Our findings imply that quantum theory admits a *stronger form of steering* than known until now, with fundamental as well as practical potential implications.

Instrumental causal networks are one of the main tools of causal inference [1, 2]. Introduced almost a century ago [3] in the context of supply-and-demand models, they find nowadays a broad range of applications, from epidemiology and clinical trials [4, 5] to econometrics [6] and ecology [7], e.g. In fact, the instrumental causal structure is special because it is the simplest one for which the strength of causal influences can be estimated solely from observational data – i.e. without interventions – even in the presence of hidden common causes [8]. Recently, considerable effort has been put into the quantisation of the classical theory of causality [1, 2], giving rise to the so-called *quantum causal networks* [9–17]. Apart from its implications in nonlocality [18–24], the young field has brought about fascinating discoveries [10, 25, 26] and applications [27–33]. However, some important causal structures have not yet received enough attention in the quantum regime. This is the case of the instrumental one.

Partly responsible for that may be the fact [12] that equipping the common cause with entanglement is not enough to violate the usual *instrumental inequalities* [8]. Instrumental inequalities are to instrumental models what Bell inequalities [34] are to local hidden-variable ones; with the difference that instrumental models are intrinsically nonlocal, involving 1-way outcome communication. In this sense, instrumental-inequality violations certify a stronger form of nonlocality than Bell violations [2]. Remarkably, in spite of the no-go result of [12], a different class [35] of instrumental inequalities has been recently shown [36] to admit a quantum violation. Besides their fundamental relevance, the violation of

instrumental inequalities with quantum resources is also potentially interesting from an applied viewpoint, as it opens a possibility towards new types of nonlocality-based protocols without the requirement of space-like separation, a major experimental overhead to current implementations.

In turn, both instrumental [2, 8, 35, 36] and Bell [34, 37] inequalities are formulated in the device-independent (DI) scenario of untrusted measurement devices, effectively treated as black boxes with classical settings (inputs) and outcomes (outputs). The DI regime is known to be experimentally much more demanding [38] than the so-called 1-sided (1S) DI one, where one of the observable nodes is a black box while the other one a trusted apparatus with full quantum control. This is the natural framework of *steering* [39], a hybrid form of quantum nonlocality intermediate between Bell and entanglement. While, in the DI scenario, local hidden-variable models enhanced with different types of communication have a long history in the literature [19, 36, 40–49], in the 1S-DI setting only input communication has received some attention [50, 51]. In contrast, output-communication enhanced local models are totally unexplored in the 1S-DI domain.

Here, we study *1S quantum instrumental (1SQI) processes* obtained from quantizing the communication-receiving node in classical instrumental causal networks, or, equivalently, from enhancing local hidden-state models with outcome communication. We introduce *1S-DI instrumental inequalities* and *non-instrumentality witnesses*, as experimentally-friendly tools to test against 1SQI models. These natu-

rally lead to a resource-theoretic measure efficiently computable via semi-definite programming: the *robustness of non-instrumentality*. Furthermore, we show that 1S-DI instrumental inequalities can be violated with little entanglement and purity at the common-cause node and, remarkably, without outcome communication, so that the violations are due solely to the states' steering. We present an experimental demonstration in an entangled-photon platform. Finally, we prove an even stronger incompatibility between steering and 1SQI processes. Namely, that quantum steering alone is enough to attain any value of the robustness. Our findings imply that steering is a stronger quantum phenomenon than previously thought, beyond classical hidden-variable models even when equipped with output communication.

1-sided quantum and quantum-common-cause instrumental processes. We start by introducing 1SQI causal networks, shown in Fig. 1 a). Nodes X, A, and B are observable, while Λ is hidden. Node B is a quantum system with Hilbert space $\mathbb{H}_B$, whereas all four other nodes encode classical random variables. Node X takes $|X|$ possible values $x \in \mathbb{Z}_X$, with the short-hand notation $\mathbb{Z}_X := \{0, \ldots |X| - 1\}$ introduced, A takes $|A|$ values $a \in \mathbb{Z}_A$, and Λ can – w.l.o.g. – be assumed to take $|\Lambda| = |A|^{|X|}$ possible values $\lambda \in \mathbb{Z}_\Lambda$ (see App. A). A 1SQI causal model assigns a probability distribution $\boldsymbol{P}_\Lambda := \{P_\lambda\}_\lambda$ to Λ, a conditional distribution $\boldsymbol{P}_{A|X,\Lambda} := \{P_{a|x,\lambda}\}_{a,x,\lambda}$ to A, and a quantum state $\varrho_{a,\lambda} \in \mathcal{B}(\mathbb{H}_B)$ to B, i.e. such that $\varrho_{a,\lambda} \geq 0$ and $\mathrm{Tr}[\varrho_{a,\lambda}] = 1$ for all a and λ. We are interested in the statistics of A and B given X. Hence, the local statistics of X is not explicitly considered here. We refer to the users at nodes A and B as Alice and Bob, respectively.

Since A is classical and B quantum, they are most-conveniently described jointly by subnormalised conditional states $\sigma_{a|x}$, which encapsulate both the probability $P_{a|x} := \mathrm{Tr}\left[\sigma_{a|x}\right]$ of a given x for Alice and the conditional state $\varrho_{a,x} := \sigma_{a|x}/P_{a|x}$ given a and x for Bob. 1SQI models produce ensembles $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{inst})} := \left\{\sigma_{a|x}^{(\mathrm{inst})} \in \mathcal{B}(\mathbb{H}_B)\right\}_{a,x}$, with

$$\sigma_{a|x}^{(\mathrm{inst})} = \sum_\lambda P_\lambda P_{a|x,\lambda} \varrho_{a,\lambda}. \tag{1}$$

We refer to $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{inst})}$ as a *1SQI assemblage*, and denote the set of all such assemblages by 1SQI. The term "assemblage" is native of the steering literature [52]. Its use here is not coincidental: there is a connection between 1SQI and steering. To see this, let us next introduce local hidden-state (LHS) models. These correspond to restricted 1SQI models without the causal influence from A to B. An *LHS as-*
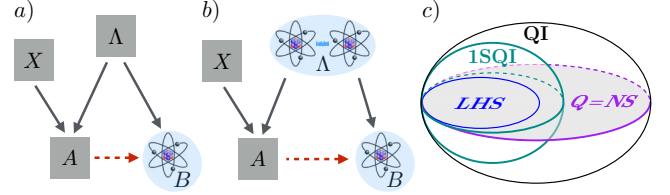

Figure 1. a) and b) Hybrid (classical-quantum) instrumental processes. Causal structures are specified by directed acyclic graphs (DAGs). Each node encodes either a classical random variable or a quantum system, and each directed edge a causal influence. The term acyclic refers to the physical requirement of no causal loops. a) 1-sided quantum instrumental (1SQI) causal networks generalise the classical instrumental causal structure to the case where node B is a quantum system. Removing the red dashed edge, in turn, leads to local hidden-state (LHS) models. b) Quantising also Λ gives what we here refer to as quantum instrumental (QI) processes: Λ is now a bipartite system in a possibly entangled quantum state, with one subsystem causally influencing A and the other one B. Removing the red dashed edge here leads to quantum (Q) models. c) Pictorial representation of the inner geometry of the set QI. 1SQI is a strict subset of QI. The lower dimensional manifold NS of non-signalling assemblages coincides (for the bipartite case) with Q, in turn containing LHS. LHS and Q are the sets studied in steering theory. Surprisingly, Q is not contained in 1SQI. What is more, for any assemblage in QI and outside of 1SQI, there is an assemblage in Q as far away from 1SQI as the former.

semblage $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{lhs})} := \left\{\sigma_{a|x}^{(\mathrm{lhs})} \in \mathcal{B}(\mathbb{H}_B)\right\}_{a,x}$ has components $\sigma_{a|x}^{(\mathrm{lhs})} := \sum_\lambda P_\lambda P_{a|x,\lambda} \varrho_\lambda$, i.e. as Eq. (1) but with $\varrho_{a,\lambda}$ independent of a. We denote the set of all LHS assemblages by LHS, and call any assemblage $\boldsymbol{\Sigma}_{A|X}$ *steerable* if $\boldsymbol{\Sigma}_{A|X} \notin \mathsf{LHS}$. Clearly, $\mathsf{LHS} \subseteq \mathsf{1SQI}$.

In turn, allowing not only for a quantum B but also for a quantum Λ defines the set QI of *quantum instrumental (QI) assemblages* $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{qinst})} := \left\{\sigma_{a|x}^{(\mathrm{qinst})} \in \mathcal{B}(\mathbb{H}_B)\right\}_{a,x}$. More precisely, we allocate to Λ a composite Hilbert space $\mathbb{H}_\Lambda = \mathbb{H}_{\Lambda_A} \otimes \mathbb{H}_{\Lambda_B}$, such that subsystems Λ_A and Λ_B causally influence nodes A and B, respectively [see Fig. 1 b)]. Hence, Λ is now a *quantum common cause* [15, 16] for A and B [53]. Accordingly, for Λ in a state $\varrho_\Lambda \in \mathcal{B}(\mathbb{H}_\Lambda)$, the resulting QI conditional states are

$$\sigma_{a|x}^{(\mathrm{qinst})} := \mathcal{E}_{B|a,\Lambda_B}\left(\mathrm{Tr}_{\Lambda_A}\left[M_x^{(a)} \otimes \mathbb{1}_{\Lambda_B} \varrho_\Lambda\right]\right). \tag{2}$$

Here, $\mathcal{E}_{B|a,\Lambda_B} : \mathcal{B}(\mathbb{H}_{\Lambda_B}) \to \mathcal{B}(\mathbb{H}_B)$ is an a-dependent completely-positive trace-preserving map and $M_x^{(a)}$ is the a-th element of an x-dependent measurement $\boldsymbol{M}_x :=$

$\left\{ M_x^{(a)} \in \mathcal{B}(\mathbb{H}_{\Lambda_A}) : M_x^{(a)} \geq 0, \sum_{a'} M_x^{(a')} = \mathbb{1}_{\Lambda_A} \right\}_a$, with $\mathbb{1}_{\Lambda_A}$ the identity on $\mathbb{H}_{\Lambda_A}$. Clearly, 1SQI $\subseteq$ QI, as Eq. (2) reduces to Eq. (1) for the specific case of ϱ_Λ separable.

On the other hand, for the particular case of $\mathcal{E}_{B|a,\Lambda_B}$ being the identity map for all a (no causal influence from A to B), QI reduces to the set Q of *quantum assemblages* $\boldsymbol{\Sigma}_{A|X}^{(q)} := \left\{ \sigma_{a|x}^{(q)} \in \mathcal{B}(\mathbb{H}_B) \right\}_{a,x}$, of components $\sigma_{a|x}^{(q)} := \mathrm{Tr}_{\Lambda_A} \left[M_x^{(a)} \otimes \mathbb{1}_{\Lambda_B} \varrho_\Lambda \right]$. Hence, Q is to QI what LHS is to 1SQI. Clearly, Q $\subseteq$ QI. In addition, from steering theory, we know that LHS $\subset$ Q. On the contrary, it holds that 1SQI $\not\subset$ Q, as Q assemblages are non-signalling while 1SQI ones not. An assemblage $\boldsymbol{\Sigma}_{A|X}$ is said to be non-signalling if

$$\varrho_B := \sum_a \sigma_{a|x} \qquad (3)$$

(the reduced state of Bob) is independent of x. Also due to non-signalling, it follows that, actually, LHS $\subset$ 1SQI and Q $\subset$ QI. We call the set of non-signalling assemblages NS. For the bipartite case under consideration, it is known that Q $=$ NS [54]. In contrast, for instrumental causal models, Bob's state can depend on x even after summing a out.

1S-DI instrumental inequalities, witnesses, and robustness. Since 1SQI is convex, any $\boldsymbol{\Sigma}_{A|X} \notin$ 1SQI is separated from 1SQI by a hyperplane, represented by an assemblage-like object $\boldsymbol{W}_{A|X} := \left\{ W_{a|x} \in \mathcal{B}(\mathbb{H}_B) \right\}_{a,x}$, with $W_{a|x}$ Hermitian, of fixed *scale* $s := \sum_{a,x} \mathrm{Tr} \left[W_{a|x} \right]$, such that

$$\left\langle \boldsymbol{W}_{A|X}, \boldsymbol{\Sigma}_{A|X}^{(\mathrm{inst})} \right\rangle := \sum_{a,x} \mathrm{Tr} \left[W_{a|x} \sigma_{a|x}^{(\mathrm{inst})} \right] \leq \beta, \qquad (4)$$

for all $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{inst})} \in$ 1SQI, and $\left\langle \boldsymbol{W}_{A|X}, \boldsymbol{\Sigma}_{A|X} \right\rangle > \beta \in \mathbb{R}$. We refer to Eq. (4) as a 1S quantum instrumental inequality with *1SQI bound* β (which depends solely on $\boldsymbol{W}_{A|X}$). Thus, $\boldsymbol{W}_{A|X}$ plays a role analogous to the normal vector of a plane in Euclidean space. We refer to $\boldsymbol{W}_{A|X}$ as a *non-instrumentality witness*. The separation is then quantified by the violation $\left\langle \boldsymbol{W}_{A|X}, \boldsymbol{\Sigma}_{A|X} \right\rangle - \beta$. Finally, we say that $\boldsymbol{W}_{A|X}$ is an *optimal non-instrumentality witness* for $\boldsymbol{\Sigma}_{A|X} \notin$ 1SQI if $\left\langle \boldsymbol{W}_{A|X}, \boldsymbol{\Sigma}_{A|X} \right\rangle - \beta \geq \left\langle \boldsymbol{W}'_{A|X}, \boldsymbol{\Sigma}_{A|X} \right\rangle - \beta'$ for all non-instrumentality witnesses $\boldsymbol{W}'_{A|X}$ with 1SQI bound β' and scale $\sum_{a,x} \mathrm{Tr} \left[W'_{a|x} \right] = s$. Remarkably, as shown in App. A, the optimal witness is obtained efficiently via semi-definite programming [52].

Witnesses are, in turn, connected with robustness measures [56, 57]. Here, we consider the *robustness of non-*
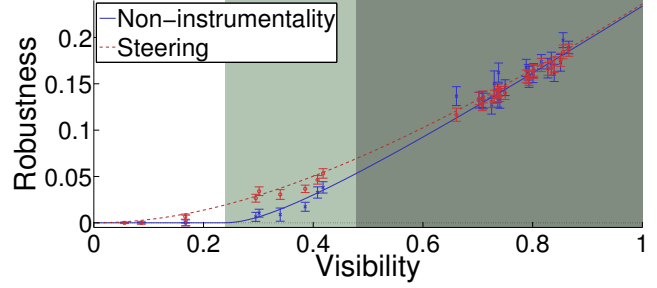

Figure 2. Robustnesses of steering and of non-instrumentality. The curves and points correspond respectively to theory and experiment. An assemblage is produced by 3 projective local measurements by Alice on the $\varrho_\Lambda = V|\Phi^+\rangle\langle\Phi^+| + \frac{(1-V)}{2}(|00\rangle\langle00| + |11\rangle\langle11|)$, generated by local dephasing on the maximally entangled state $|\Phi^+\rangle = (1/\sqrt{2})(|00\rangle + |11\rangle)$ with strength given by the visibility V. The assemblage is steerable in the usual sense for all $V > 0$, and is compatible with 1SQI models up to $V < 0.24$ (white region). That point is the onset of a stronger form of quantum steering than explainable by 1SQI models (light and dark green). Finally, for $V > (9 - 4\sqrt{2})/7 \simeq 0.48$ (dark green), the assemblage produces (under the optimal [55] a-dependent measurements on Bob's qubit) black-box correlations that violate the DI instrumental inequality of [35, 36] (see App. C). That is, the assemblage's incompatibility with 1SQI can be verified in the 1S DI setting for a broader range of visibilities than in the fully DI one. Error bars obtained assuming Poissonian distributions for the photon counts.

instrumentality R_{ni}, defined, for any $\boldsymbol{\Sigma}_{A|X} \in$ QI, as

$$R_{\mathrm{ni}}(\boldsymbol{\Sigma}_{A|X}) := \min \left\{ t \in \mathbb{R}_{\geq 0} : \frac{\boldsymbol{\Sigma}_{A|X} + t\,\boldsymbol{\Pi}_{A|X}}{1+t} \in \mathrm{1SQI} \right\}. \qquad (5)$$

It measures the minimal mixing $t/(1+t)$ with any $\boldsymbol{\Pi}_{A|X} \in$ 1SQI that $\boldsymbol{\Sigma}_{A|X}$ tolerates before the mixture enters 1SQI. Interestingly, R_{ni} is a measure of non-instrumentality in the formal, resource-theoretic sense [58–60], as we explicitly show in App. D. Moreover, we note that other choices of "noise" types are possible, giving rise to different variants of R_{ni}. However, the choice $\boldsymbol{\Pi}_{A|X} \in$ 1SQI is particularly convenient as it yields the robustness efficiently computable through a semi-definite programming optimisation. In fact, such optimisation shows that $R_{\mathrm{ni}}(\boldsymbol{\Sigma}_{A|X}) = \left\langle \boldsymbol{V}_{A|X}, \boldsymbol{\Sigma}_{A|X} \right\rangle - 1$, where $\boldsymbol{V}_{A|X}$ is the optimal for $\boldsymbol{\Sigma}_{A|X}$ over a simple subclass of non-instrumentality witnesses (see App. E for details). We call $\boldsymbol{V}_{A|X}$ the *optimal robustness witness* for $\boldsymbol{\Sigma}_{A|X}$.

Fig. 2 shows R_{ni}, together with the usual steering robustness [52], for Q assemblages obtained from local measure-

ments (by Alice) on states with different degrees of entanglement and purity. Some assemblages in the figure have positive steering robustness and $R_{\mathrm{ni}} = 0$. This confirms that LHS $\subset$ 1SQI: Alice's outcome signalling indeed provides the models with more descriptive power. However, the figure also shows assemblages with $R_{\mathrm{ni}} > 0$. This implies the following theorem, proven also analytically in App. B.

Theorem 1 (Q $\nsubseteq$ 1SQI)**.** *Outcome signalling is not enough for LHS models to reproduce quantum steering.*

It is instructive to compare with the fully DI case, where Bob's instrument is also a black box. There, usual Bell-like correlations (where Alice and Bob have independent inputs) obtained without outcome communication are known to be stronger than local hidden-variable models augmented with outcome signalling from Alice [19, 48, 61]. However, for DI instrumental processes, in contrast, if Bob does not actively exploit Alice's output his measurement setting is fixed [62]. Hence, no matter how entangled ϱ_Λ, or what measurements Alice makes, the resulting correlations will be automatically compatible with local hidden-variable models with no input for Bob, a subclass of classical instrumental models. In other words, if Bob applies a measurement that does not depend on a, the correlations trivially fulfill any DI instrumental inequality, including the recent one of Ref. [36]. Thus, incompatibility with the instrumental DAG without output signalling is a distinctive feature of the 1-sided DI case.

Finally, R_{ni} could in principle attain higher values over QI than over Q. After all, the former allows for signalling while the latter does not. Surprisingly, this is false. The following theorem, proven in App. F, holds instead.

Theorem 2 ($R_{\mathrm{ni}}(\mathsf{Q}) = R_{\mathrm{ni}}(\mathsf{QI})$)**.** *For every quantum instrumental assemblage, there exists a Q assemblage with the same non-instrumentality robustness.*

From a practical viewpoint, the theorem provides a significant computational shortcut in the task of, given a fixed value of R_{ni}, finding an assemblage with that robustness. Because the theorem allows one to restrict the search to quantum assemblages, instead of searching over all QI ones [63]. From a fundamental perspective, in turn, it has implications in the inner geometry of QI. Namely, it tells us that, for any point $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{qinst})} \in \mathsf{QI} \setminus \mathsf{1SQI}$, there is always a point $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{q})} \in \mathsf{Q}$ as far away from 1SQI as $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{qinst})}$. This does not contradict the fact that Q $\subset$ QI, because Q is a lower-dimensional manifold (the NS one) of QI. Theorem 2 thus suggests that $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{q})}$ is a kind of projection of $\boldsymbol{\Sigma}_{A|X}^{(\mathrm{qinst})}$ onto NS [see Fig. 1 c)].
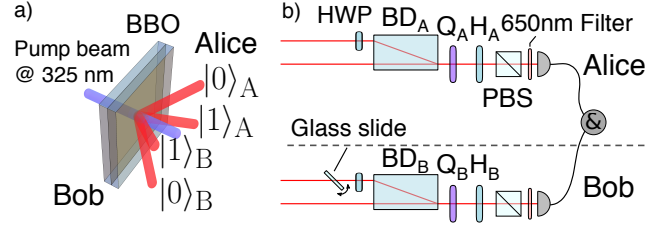

Figure 3. Experimental setup. (a) The BBO crystal is pumped by a vertically polarized laser beam at 325 nm from a He-Cd source, so that a pair of down-converted photons, at 650 nm, emerge in an entangled state of momenta due to momentum conservation. The momenta encode the two common-cause qubits. All but two momentum modes, corresponding to $|00\rangle$ and $|11\rangle$, are filtered out. The produced momentum state is the two-qubit maximally entangled state $|\Phi^+\rangle\langle\Phi^+|$. One qubit is sent to Alice and the other one to Bob. (b) Side view of each party's device, depicting the preparation (by Alice) and measurement (by Bob) of the assemblage. Beam displacers BD_A and BD_B and half-wave plates (HWP) map the momenta of each photon into its polarisations. A glass slide introduces dephasing between Bob's two paths, allowing us to tune the visibility V of the resulting dephased state ϱ_Λ from $V \sim 0$ to $V \sim 94,5\%$. Quarter-wave plates (Q_A and Q_B), half-wave plates (H_A and H_B), polarised beam splitters (PBS), and a coincidence-photon detector implement projective polarization measurements.

Experimental demonstration. We verified the gap between quantum steering and non-instrumentality using entangled-photon pairs produced by spontaneous parametric down conversions in a BBO crystal [64–66] (see Fig. 3). Alice's device is taken as the black box, so her measurement settings (wave-plate angles) and outcomes encode the bits x and a, respectively. Each measurement by her accounts for the preparation of the assemblage element $\sigma_{a|x}^{(\mathrm{q})}$. In contrast, Bob's measurements allow for state tomography of each $\sigma_{a|x}^{(\mathrm{q})}$. We produce assemblages of different purities and steering by applying local dephasing of different strengths V on Bob's qubit and having Alice measure the observables $-(\sigma_x + \sigma_z)/\sqrt{2}$, σ_x, or σ_z, where σ_x and σ_z are respectively the first and third Pauli matrices. These three observables correspond to the input choices $x = 0$, 1, and 2, respectively. We tomographically reconstruct each produced assemblage and evaluate R_{ni}, whose values are those displayed in Fig. 2.

Final discussion. 1SQI processes are a generalisation of both classical instrumental processes (native of causal inference [1, 2]) to the case where the final node is quantum and of local hidden-state models (native of steering theory [39, 52]),

thus unifying two previously disconnected topics. We introduced inequalities and witnesses to test against such 1SQI causal models. These can, in particular, detect entanglement with a single trusted device even in the presence of 1-way outcome signalling. Strikingly, they can also be violated with entanglement alone, i.e. without using outcome communication. Hence, outcome communication is not enough to explain quantum steering. Interestingly, this is a distinctive feature of the 1S DI setting: For DI instrumental processes, if Bob's setting is independent of Alice's outcome, any state produces local hidden-variable correlations, automatically compatible with classical instrumental models.

The proposed quantifier (robustness) of non-instrumentality is efficiently computable via semi-definite programming. Moreover, we prove in the appendix that it is a formal, resource-theoretic monotone. With it, we showed an even stronger incompatibility between steering and 1SQI processes: that any value of the robustness can be attained with steering alone. We experimentally verified our predictions in an entangled-photon platform. The experiment is simple, but proves that quantum states can be steerable in a stronger way than previously reported.

Finally, the fact that quantum mechanics allows one to falsify – with quantum control only at a single lab – classical explanations even when these exploit output signalling is not only relevant from the perspectives of quantum foundations and causal inference but also promising from an applied one. More precisely, quantum-nonlocality applications are experimentally less demanding in the 1S DI regime than the fully DI one, as already mentioned. To this, our findings now add the possibility of steering-based protocols with the additional experimentally-appealing feature of no need for space-like separation. We note that, even if B is in the future light-cone of A (and, therefore, also of X), direct causal influences from X to B can be ruled out (and so an underlying instrumental causal structure guaranteed) with interventions on A, e.g. There, cryptographic or randomess-generation protocols based on 1S quantum instrumental inequality violations are conceivable. It would thus be interesting to explore steering beyond outcome signalling as a potential resource for information processing in comparison with conventional steering-based schemes [38] requiring space-like separation. Our results open new venues for research in that direction.

ACKNOWLEDGMENTS

[1] P. Spirtes, N. Glymour, and R. Scheienes, *Causation, Prediction, and Search, 2nd ed.* (The MIT Press, 2001).

[2] J. Pearl, *Causality* (Cambridge University Press, 2009).

[3] P. G. Wright *et al.*, *Tariff on animal and vegetable oils* (The Macmillan Co., 1928).

[4] A. Balke and J. Pearl, "Bounds on Treatment Effects From Studies With Imperfect Compliance," Journal of the American statistical Association **92**, 1171 (1997).

[5] S. Greenland, "An introduction to instrumental variables for epidemiologists," International Journal of Epidemiology **29**, 722 (2000).

[6] J. D. Angrist, G. W. Imbens, and D. B. Rubin, "Identification of causal effects using instrumental variables," Journal of the American statistical Association **91**, 444–455 (1996).

[7] S. Creel and M. Creel, "Density dependence and climate effects in rocky mountain elk: an application of regression with instrumental variables for population time series with sampling error," Journal of Animal Ecology **78**, 1291–1297 (2009).

[8] J. Pearl, "On the testability of causal models with latent and in-

strumental variables," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence* (Morgan Kaufmann Publishers Inc., 1995) pp. 435–443.

[9] G. Chiribella, G. M. D'Ariano, and P. Perinotti, "Theoretical framework for quantum networks," Phys. Rev. A **80**, 022339 (2009).

[10] O. Oreshkov, F. Costa, and C. Brukner, "Quantum correlations with no causal order," Nature Communications **3**, 1092 (2012).

[11] M. S. Leifer and R. W. Spekkens, "Towards a formulation of quantum theory as a causally neutral theory of bayesian inference," Phys. Rev. A **88**, 052130 (2013).

[12] J. Henson, R. Lal, and M. F. Pusey, "Theory-independent limits on correlations from generalized bayesian networks," New J. Phys. **16**, 113043 (2014).

[13] R. Chaves, C. Majenz, and D. Gross, "Information–theoretic implications of quantum causal structures," Nat. Commun. **6**, 5766 (2015).

[14] J. Pienaar and C. Brukner, "A graph-separation theorem for quantum causal models," New J. Phys. **17**, 073020 (2015).

[15] F. Costa and S. Shrapnel, "Quantum causal modelling," New Journal of Physics **18**, 063032 (2016).

[16] John-Mark A. Allen, Jonathan Barrett, Dominic C. Horsman, Ciarán M. Lee, and Robert W. Spekkens, "Quantum common causes and quantum causal models," Phys. Rev. X **7**, 031021 (2017).

[17] C. Portmann, C. Matt, U. Maurer, R. Renner, and B. Tackmann, "Causal boxes: quantum information-processing systems closed under composition," IEEE Transactions on Information Theory (2017).

[18] T. Fritz, "Beyond bell's theorem: correlation scenarios," New Journal of Physics **14**, 103001 (2012).

[19] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, "Unifying framework for relaxations of the causal assumptions in Bell's theorem," Phys. Rev. Lett. **114**, 140403 (2015).

[20] R. Chaves, "Polynomial bell inequalities," Phys. Rev. Lett. **116**, 010402 (2016).

[21] T. Fritz, "Beyond bell's theorem ii: Scenarios with arbitrary causal structure," Communications in Mathematical Physics **341**, 391–434 (2016).

[22] D. Rosset, C. Branciard, T. J. Barnea, G. Pütz, N. Brunner, and N. Gisin, "Nonlinear Bell inequalities tailored for quantum networks," Phys. Rev. Lett. **116**, 010403 (2016).

[23] E. Wolfe, R. W. Spekkens, and T. Fritz, "The inflation technique for causal inference with latent variables," arXiv preprint arXiv:1609.00672 (2016).

[24] R. Chaves, D. Cavalcanti, and L. Aolita, "Causal hierarchy of multipartite Bell nonlocality," Quantum **1**, 23 (2017).

[25] G. Chiribella, G. M. D'Ariano, P. Perinotti, and B. Valiron, "Quantum computations without definite causal structure," Phys. Rev. A, **88**, 022318, arXiv:0912.0195 (2013).

[26] C. Branciard, M. Araújo, A. Feix, F. Costa, and C. Brukner, "The simplest causal inequalities and their violation," New Journal of Physics **18**, 013008 (2015).

[27] M. Araújo, C. Branciard, F. Costa, A. Feix, C. Giarmatzi, and C. Brukner, "Witnessing causal nonseparability," New J. Phys. **17**, 102001 (2015).

[28] K. Ried, M. Agnew, L. Vermeyden, D. Janzing, R. W. Spekkens, and K. J. Resch, "A quantum advantage for inferring causal structure," Nat Phys **11**, 414–420 (2015).

[29] Rafael Chaves, Jonatan Bohr Brask, and Nicolas Brunner, "Device-independent tests of entropy," Phys. Rev. Lett. **115**, 110501 (2015).

[30] Jean-Philippe W MacLean, Katja Ried, Robert W Spekkens, and Kevin J Resch, "Quantum-coherent mixtures of causal relations," Nature Communications **8** (2017).

[31] M. Araújo, F. Costa, and C. Brukner, "Computational advantage from quantum-controlled ordering of gates," Phys. Rev. Lett. **113**, 250402 (2014).

[32] P. A. Guérin, A. Feix, M. Araújo, and C. Brukner, "Exponential communication complexity advantage from quantum superposition of the direction of communication," Phys. Rev. Lett. **117**, 100502 (2016).

[33] R. Rossi, "Restrictions for the causal inferences in an interferometric system," Phys. Rev. A **96**, 012106 (2017).

[34] J. S. Bell, "On the Einstein–Podolsky–Rosen paradox," Physics **1**, 195 (1964).

[35] B. Bonet, "Instrumentality tests revisited," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence* (Morgan Kaufmann Publishers Inc., 2001) pp. 48–55.

[36] R. Chaves, G. Carvacho, I. Agresti, L. Aolita, V. Di Giulio, S. Giacomini, and F. Sciarrino, "Quantum violation of an instrumental test," Nature Physics (2017).

[37] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell nonlocality," Rev. Mod. Phys. **86**, 419–478 (2014).

[38] C. Branciard, E. G. Cavalcanti, S. P. Walborn, V. Scarani, and H. M. Wiseman, "One-sided device-independent quantum key distribution: Security, feasibility, and the connection with steering," Phys. Rev. A **85**, 010301(R) (2012).

[39] M. D. Reid, P. D. Drummond, W. P. Bowen, E. G. Cavalcanti, P. K. Lam, H. A. Bachor, U. L. Andersen, and G. Leuchs, "Colloquium: The Einstein-Podolsky-Rosen paradox: From concepts to applications," Rev. Mod. Phys. **81**, 1727 (2009).

[40] Tim Maudlin, "Bell's inequality, information transmission, and prism models," PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association **1992**, 404–417 (1992).

[41] N. Gisin and B. Gisin, "A local hidden variable model of quantum correlation exploiting the detection loophole," Physics Letters A **260**, 323 – 327 (1999).

[42] Gilles Brassard, Richard Cleve, and Alain Tapp, "Cost of exactly simulating quantum entanglement with classical communication," Phys. Rev. Lett. **83**, 1874–1877 (1999).

[43] Michael Steiner, "Towards quantifying non-local information transfer: finite-bit non-locality," Physics Letters A **270**, 239 – 244 (2000).

[44] B. F. Toner and D. Bacon, "Communication cost of simulating bell correlations," Phys. Rev. Lett. **91**, 187904 (2003).

[45] Julien Degorre, Sophie Laplante, and Jérémie Roland, "Simulating quantum correlations as a distributed sampling problem," Phys. Rev. A **72**, 062314 (2005).

[46] Oded Regev and Ben Toner, "Simulating quantum correlations with finite communication," SIAM Journal on Computing **39**, 1562–1580 (2009), preliminary version in FOCS'07.

[47] Marcin Pawlowski, Johannes Kofler, Tomasz Paterek, Michael Seevinck, and Caslav Brukner, "Non-local setting and outcome information for violation of bell's inequality," New Journal of Physics **12**, 083051 (2010).

[48] M. Ringbauer, C. Giarmatzi, R. Chaves, F. Costa, A. G. White, and A. Fedrizzi, "Experimental test of nonlocal causality," Science Advances **2**, e1600162 (2016).

[49] J. Bohr Brask and R. Chaves, "Bell scenarios with communication," Phys. A: Math. Theor. **50**, 094001 (2017).

[50] A. B. Sainz, L. Aolita, N. Brunner, R. Gallego, and P. Skrzypczyk, "Classical communication cost of quantum

steering," Phys. Rev. A **94**, 012308 (2016).

[51] S. Nagy and T. Vértesi, "EPR Steering inequalities with Communication Assistance," Sci. Rep. **6**, 21634 (2016).

[52] D. Cavalcanti and P. Skrzypczyk, "Quantum steering: a review with focus on semidefinite programming," Reports on Progress in Physics **80**, 024001 (2016).

[53] Importantly, node A is kept classical throughout. The term "quantum" in QI is used here as a short form to refer to "quantum common-cause". Thus, QI should not be mistaken with the fully quantum case where all nodes, including A, are quantised.

[54] A. B. Sainz, N. Brunner, D. Cavalcanti, P. Skrzypczyk, and T. Vértesi, "Postquantum steering," Phys. Rev. Lett. **115**, 190403 (2015).

[55] M. M. Taddei, R. V. Nery, and L. Aolita, "Necessary and sufficient conditions for multipartite Bell violations with only one trusted device," Phys. Rev. A **94**, 032106 (2016).

[56] Guifré Vidal and Rolf Tarrach, "Robustness of entanglement," Phys. Rev. A **59**, 141–155 (1999).

[57] M. Piani and J. Watrous, "Necessary and sufficient quantum information characterization of Einstein-Podolsky-Rosen steering," Phys. Rev. Lett. **114**, 060404 (2015).

[58] R. Gallego and L. Aolita, "Resource theory of steering," Phys. Rev. X **5**, 041008 (2015).

[59] R. Gallego and L. Aolita, "Nonlocality free wirings and the distinguishability between Bell boxes," Phys. Rev. A **95**, 032118 (2017).

[60] B. Amaral, A. Cabello, M. T. Cunha, and L. Aolita, "Noncontextual wirings," arXiv: 1705.07911 (2017).

[61] M. Ringbauer and R. Chaves, "Probing the non-classicality of temporal correlations," arXiv: 1704.05469 (2017).

[62] Note that, interestingly, instrumental causal models display the same DAG for both the DI and the 1-sided DI cases. This is not the case for LHV and LHS models. The former need an extra (input) node with respect to the latter.

[63] Interestingly, optimising the violation of a fixed NI witness over all quantum assemblages can – thanks to the no-signalling constraint – be directly recast as an SDP problem. In contrast, this is not true for QI assemblages.

[64] O. J. Farías, G. H. Aguilar, A. Valdés-Hernández, P. H. Souto Ribeiro, L. Davidovich, and S. P. Walborn, "Observation of the emergence of multipartite entanglement between a bipartite system and its environment," Phys. Rev. Lett. **109**, 150403 (2012).

[65] G. H. Aguilar, O. J. Farías, A. Valdés-Hernández, P. H. Souto Ribeiro, L. Davidovich, and S. P. Walborn, "Flow of quantum correlations from a two-qubit system to its environment," Phys. Rev. A **89**, 022339 (2014).

[66] D. Cavalcanti, P. Skrzypczyk, G. H. Aguilar, R. V. Nery, P. H. Souto Ribeiro, and S. P. Walborn, "Detection of entanglement in asymmetric quantum networks and multipartite quantum steering." Nature communications **6**, 7941 (2015).

[67] Programs used to perform the analyses of the assemblages can be found at https://git.io/vbHqN.

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

Appendix A: 1-sided quantum instrumentality as a semi-definite programming membership problem

In this section, we consider the problems of how to determine if a given arbitrary assemblage $\mathbf{\Sigma}_{A|X}$ is or not in 1SQI and how to determine its optimal non-instrumentality witness, together with its corresponding violation. As in the membership problem for LHS and the determination of optimal steering witnesses of standard steering theory [52], these problems turn out to admit a formulation as a semi-definite programe (SDP). SDPs deal with optimisations of a linear objective function over a matrix space defined by linear and positive-semidefinite constraints. Because of this, SDPs are exact in the sense that the solutions they return are guaranteed not to get stuck at local maxima or minima [68].

To this end, we express the conditional states in Eq. (1) as

$$\sigma_{a|x}^{(\text{inst})} = \sum_{\tilde{\lambda}} D_{\tilde{\lambda}}(a|x) \, \tilde{\sigma}_{a,\tilde{\lambda}}, \tag{A1}$$

where $D_{\tilde{\lambda}}$ is the $\tilde{\lambda}$-th deterministic response function and $\tilde{\sigma}_{a,\tilde{\lambda}} := \sum_{\lambda} \tilde{P}_{\tilde{\lambda}|\lambda} \, P_\lambda \, \varrho_{a,\lambda}$, with $\tilde{P}_{\tilde{\lambda}|\lambda}$ defined such that $P_{a|x,\lambda} =: \sum_{\tilde{\lambda}} \tilde{P}_{\tilde{\lambda}|\lambda} \, D_{\tilde{\lambda}}(a|x)$. There are as many such functions as hidden-variable values, i.e. $|\tilde{\Lambda}| = |\Lambda| = |A|^{|X|}$. In addition, the conditional states $\tilde{\sigma}_{a,\tilde{\lambda}}$ are subnormalised such that

$$\text{Tr}\left[\tilde{\sigma}_{a,\tilde{\lambda}}\right] = \sum_{\lambda} \tilde{P}_{\tilde{\lambda}|\lambda} \, P_\lambda =: \tilde{P}_{\tilde{\lambda}}, \tag{A2}$$

for all $a \in \mathbb{Z}_A$ and all $\tilde{\lambda} \in \mathbb{Z}_\Lambda$, i.e. their trace is independent of a. Note that the distribution $\tilde{\mathbf{P}}_\Lambda := \{\tilde{P}_{\tilde{\lambda}}\}_{\tilde{\lambda}}$ is automatically normalized if so is $\mathbf{\Sigma}_{A|X}$.

We can then recast the membership problem of $\mathbf{\Sigma}_{A|X}$ for 1SQI, i.e. whether $\mathbf{\Sigma}_{A|X}$ admits or not a decomposition as in Eq. (1), directly as an SDP feasibility test. This can be conveniently expressed by the following optimisation.

$$\text{Given} \quad \mathbf{\Sigma}_{A|X} = \left\{\sigma_{a|x}\right\}_{a,x},$$

$$\min_{\{\tilde{\sigma}_{a,\lambda} \in \mathcal{B}(\mathbb{H}_B)\}_{a,\lambda}} \quad \mu, \tag{A3a}$$

$$\text{s. t.} \quad \sigma_{a|x} = \sum_\lambda D_\lambda(a|x) \, \tilde{\sigma}_{a,\lambda}, \tag{A3b}$$

$$\text{with} \quad \text{Tr}\left[\tilde{\sigma}_{a,\lambda}\right] = \text{Tr}\left[\tilde{\sigma}_{0,\lambda}\right], \tag{A3c}$$

$$\text{and} \quad \tilde{\sigma}_{a,\lambda} \geq -\mu \, \mathbb{1}_B. \tag{A3d}$$

Eqs. (A3b) and (A3c) encode the constraints in Eqs. (A1) and (A2), respectively. Hence, the minimisation in Eq. (A3a)

amounts to finding an 1SQI decomposition in terms of conditional states $\tilde{\sigma}_{a,\tilde{\lambda}}$ as positive as possible, in the sense of satisfying the constraint of Eq. (A3d) with μ as negative as possible. When the objective function μ reaches a non-positive value, a decomposition as in Eq. (A1) is feasible with some $\tilde{\sigma}_{a,\tilde{\lambda}} \geq 0$, and vice versa. That is, any value $\mu > 0$ returned by the optimisation is equivalent to an 1SQI-decomposition being infeasible for $\mathbf{\Sigma}_{A|X}$, i.e. to $\mathbf{\Sigma}_{A|X} \notin$ 1SQI.

By virtue of the duality theory of semi-definite programming [52, 68], every such SDP admits a dual, equivalent formulation as follows.

$$\text{Given} \quad \mathbf{\Sigma}_{A|X} = \left\{\sigma_{a|x}\right\}_{a,x},$$

$$\max_{\{W_{a|x} \in \mathcal{B}(\mathbb{H}_B)\}_{a,x}} \quad \sum_{a,x} \text{Tr}\left[W_{a|x} \, \sigma_{a|x}\right], \tag{A4a}$$

$$\text{s. t.} \quad \sum_x W_{a|x} \, D_\lambda(a|x) \leq C_{a,\lambda} \, \mathbb{1}_B, \tag{A4b}$$

$$\text{with} \quad \sum_a C_{a,\lambda} = 0, \tag{A4c}$$

$$\text{and} \quad \sum_{a,x,\lambda} \text{Tr}\left[W_{a|x} \, D_\lambda(a|x)\right] = -1. \tag{A4d}$$

where $C_{a,\lambda} \in \mathbb{R}$ for all $a \in \mathbb{Z}_A$ and all $\lambda \in \mathbb{Z}_\Lambda$. Eqs. (A4b) and (A4c) together imply that $\sum_{a,x} \text{Tr}\left[W_{a|x} \sum_{\tilde{\lambda}} D_{\tilde{\lambda}}(a|x) \, \tilde{\sigma}_{a,\tilde{\lambda}}\right] \leq 0$, for any conditional states $\tilde{\sigma}_{a,\tilde{\lambda}}$ satisfying Eq. (A2). So, these two equations encode the constraints that the assemblage-like object $\mathbf{W}_{A|X} = \{W_{a|x} \in \mathcal{B}(\mathbb{H}_B)\}_{a,x}$ of Hermitian operators $W_{a|x}$ returned by the optimisation is a non-instrumentality witness for some $\beta \geq 0$. Eq. (A4d) fixes the scale of $\mathbf{W}_{A|X}$, which prevents the maximisation in Eq. (A4a) from diverging to ∞. Indeed, using the fact that $\sum_\lambda D_\lambda(a|x) = |A|^{|X|-1}$, Eq. (A4d) yields

$$s = -\frac{1}{|A|^{|X|-1}}. \tag{A5}$$

Other choices of scaling are valid, but they must be accompanied by a corresponding rescaling factor for the primal objective function in Eq. (A3a). Finally, the maximisation in Eq. (A4a) guarantees that $\mathbf{W}_{A|X}$ is the optimal non-instrumentality witness for $\mathbf{\Sigma}_{A|X}$ and that it is therefore (due to the convexity of 1SQI) tight – i.e. that $\beta = 0$ –, as we wanted to show. In other words, the maximisation returns a positive value if, and only if, $\mathbf{\Sigma}_{A|X} \notin$ 1SQI.

Finally, it is important to mention that the primal and dual SDPs, given respectively by Eqs. (A3) and (A4), satisfy a

convenient property called *strong duality* [52, 68]. By virtue of this, the primal and dual objective functions, i.e. μ and $\sum_{a,x} \text{Tr} \left[W_{a|x} \, \sigma_{a|x} \right]$, respectively, converge to the same optimal values. That is, the minimum of Eq. (A3a) and the maximum of Eq. (A4a) are guaranteed to coincide.

Appendix B: Proof of theorem 1

Here, we analytically prove theorem 1. We give a non-instrumentality witness $\boldsymbol{W}_{A|X}$ and a quantum assemblage $\boldsymbol{\Sigma}_{A|X}^{(q)} \in \mathsf{Q}$ such that $\boldsymbol{\Sigma}_{A|X}^{(q)}$ violates the 1S quantum instrumental inequality defined by $\boldsymbol{W}_{A|X}$. The assemblage we use is the same as that of Fig. 2 in the main text for $V = 1$ (no dephasing), i.e. the one obtained from $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ through projective local measurements by Alice in the bases $-(\sigma_x + \sigma_z)/\sqrt{2}$, σ_x, and σ_z. As our witness, we use the assemblage itself multiplied by a factor 2 for normalization purposes, i.e. $\boldsymbol{W}_{A|X} = 2\,\boldsymbol{\Sigma}_{A|X}^{(q)}$. Below, we show that the 1SQI bound corresponding to this witness is $\beta = 2 + 1/\sqrt{2}$. On the other hand, it is immediate to see that

$$\langle \boldsymbol{W}_{A|X}, \, \boldsymbol{\Sigma}_{A|X}^{(q)} \rangle = 3 > 2 + 1/\sqrt{2}. \tag{B1}$$

This implies that $\boldsymbol{\Sigma}_{A|X}^{(q)} \notin \mathsf{1SQI}$ and, therefore, that $\mathsf{Q} \not\subseteq \mathsf{1SQI}$.

To prove that $\beta = 2 + 1/\sqrt{2}$, we analytically maximise $\langle \boldsymbol{W}_{A|X}, \, \boldsymbol{\Sigma}_{A|X}^{(\text{inst})} \rangle$ over all $\boldsymbol{\Sigma}_{A|X}^{(\text{inst})} \in \mathsf{1SQI}$ and obtain the claimed maximum $2 + 1/\sqrt{2}$. To this end, note first that the components of the quantum assemblage $\boldsymbol{\Sigma}_{A|X}^{(q)}$ in question are rank 1, so that $\text{Tr}\left[W_{a,x} \, W_{a',x} \right] = \delta_{a,a'}$. Then, using that $\sigma_{a|x}^{(\text{inst})} = \sum_{\lambda \in \Lambda} P(\lambda)\, D_\lambda(a|x)\, \rho_{a,\lambda}$, one gets

$$\langle \boldsymbol{W}_{A|X}, \, \boldsymbol{\Sigma}_{A|X}^{(\text{inst})} \rangle = \frac{3}{2} + \frac{1}{2} P(\lambda) \left(\boldsymbol{v}_{0,\lambda} \cdot \boldsymbol{r}_{0,\lambda} - \boldsymbol{v}_{1,\lambda} \cdot \boldsymbol{r}_{1,\lambda} \right), \tag{B2}$$

where $\boldsymbol{r}_{a,\lambda} := \sum_{j=1}^{3} \text{Tr}[\rho_{a,\lambda}\, \sigma_j]\, \boldsymbol{e}_j$ and

$$\boldsymbol{v}_{a,\lambda} := \sum_{x=0}^{2} D_\lambda(a|x)\, \boldsymbol{u}_x, \tag{B3}$$

with $\boldsymbol{u}_x := \sum_{j=1}^{3} \text{Tr}[W_{0,x}\, \sigma_j]\, \boldsymbol{e}_j$ and $\boldsymbol{e}_j$ the j-th element of the canonical orthogonal basis of $\mathbb{R}^3$, for $j = 1, 2,$ or 3. Finally, optimising over $P(\lambda)$ and $\boldsymbol{r}_{a,\lambda}$ yields

$$\langle \boldsymbol{W}_{A|X}, \, \boldsymbol{\Sigma}_{A|X}^{(\text{inst})} \rangle = \frac{3}{2} + \frac{1}{2} \max_{\lambda \in \Lambda} \|\boldsymbol{v}_{0,\lambda}\| + \|\boldsymbol{v}_{1,\lambda}\|. \tag{B4}$$

For the vectors $\boldsymbol{u}_x$ used here (namely $\boldsymbol{u}_0 = -(\boldsymbol{e}_1 + \boldsymbol{e}_3)/\sqrt{2}$, $\boldsymbol{u}_1 = \boldsymbol{e}_1$, and $\boldsymbol{u}_2 = \boldsymbol{e}_3$) we obtain $\max_{\lambda \in \Lambda} \|\boldsymbol{v}_{0,\lambda}\| + \|\boldsymbol{v}_{1,\lambda}\| = 1 + \sqrt{2}$, from which the value $\beta = 2 + 1/\sqrt{2}$ follows and the proof is completed.

Appendix C: Device-independent instrumental inequality

We use the linear inequality derived in [35] and recently revisited in [36] to test for violations of the classical instrumental model by our assemblages, under an a-dependent measurement M_a by Bob (his input is equal to the output of Alice's black box). The inequality can be expressed as

$$\sum_{a,\, b \in \{0,1\}} \left[(-1)^a - (-1)^b - (-1)^{a+b} \right]\, P(a,b|0)$$
$$+ 2\,(-1)^b\, P(a,b|1) + 2(-1)^{a+b}\, P(a,b|2) \leq 3, \tag{C1}$$

where $P(a,b|x) := \text{Tr}\left[M_a^{(b)}\, \sigma_{a|x} \right]$, with $M_a^{(b)}$ the b-th element of the a-th measurement of Bob's. The optimal measurements by Bob for the maximal violation are obtained through the analytical technique of Ref. [55].

Appendix D: Monotonicity of R_{ni} under free operations of non-instrumentality

In this section we prove that R_{ni} is a *non-instrumentality monotone* for any linear class of free operations of non-instrumentality. We leave the details of the resource theory of non-instrumentality (in particular the explicit form of the corresponding free operations) for future work and prove monotonicity solely from the abstract generic properties of free operations. That is, we prove that R_{ni} is monotonous (non-increasing) under any linear map satisfying the essential free-operation requirement that $\mathcal{M}(\boldsymbol{\Sigma}_{A|X}) \in \mathsf{1SQI}$ for all $\boldsymbol{\Sigma}_{A|X} \in \mathsf{1SQI}$. The proof is similar to that [58] of steering monotonicity for the steering robustness [52].

By definition, $R_{\text{ni}}(\boldsymbol{\Sigma}_{A|X})$ is the minimal value $t^* \in \mathbb{R}_{\geq 0}$ such that

$$\boldsymbol{\Sigma}_{A|X} + t^*\, \boldsymbol{\Pi}_{A|X} = (1 + t^*)\, \boldsymbol{\Sigma}_{A|X}^{(\text{inst})}, \tag{D1}$$

for some $\boldsymbol{\Pi}_{A|X} \in \mathsf{1SQI}$ and $\boldsymbol{\Sigma}_{A|X}^{(\text{inst})} \in \mathsf{1SQI}$. Applying $\mathcal{M}$ to both sides of this equation gives

$$\mathcal{M}(\boldsymbol{\Sigma}_{A|X}) + t^*\, \mathcal{M}(\boldsymbol{\Pi}_{A|X}) = (1 + t^*)\, \mathcal{M}\left(\boldsymbol{\Sigma}_{A|X}^{(\text{inst})} \right), \tag{D2}$$

where the linearity of $\mathcal{M}$ has been used. Now, since $\mathcal{M}(\mathbf{\Pi}_{A|X})$, $\mathcal{M}(\mathbf{\Sigma}_{A|X}^{(\text{inst})}) \in$ 1SQI (because $\mathcal{M}$ is a free operation of QI), Eq. (D2) realises a particular decomposition for $\mathcal{M}(\mathbf{\Sigma}_{A|X})$ of the form of that of Eq. (D1) for $\mathbf{\Sigma}_{A|X}$. Thus, t^* must necessarily be larger or equal than the corresponding minimum for $\mathcal{M}(\mathbf{\Sigma}_{A|X})$. That is,

$$R_{\text{ni}}(\mathcal{M}(\mathbf{\Sigma}_{A|X})) \leq R_{\text{ni}}(\mathbf{\Sigma}_{A|X}), \tag{D3}$$

which proves that R_{ni} is a non-instrumentality monotone.

Appendix E: Robustness of non-instrumentality as an SDP optimisation

Eq. (5) can be re-expressed as $R_{\text{ni}}(\mathbf{\Sigma}_{A|X}) = t^*$, with t^* defined by Eq. (D1). This implies that

$$\sigma_{a|x} = (1 + t^*) \sum_{\tilde{\lambda}} D_{\tilde{\lambda}}(a|x)\, \tilde{\sigma}_{a,\tilde{\lambda}} - t^* \pi_{a|x}, \tag{E1}$$

with $\mathbf{\Pi}_{A|X} = \{\pi_{a|x} \geq 0\}_{a,x} \in$ 1SQI. Hence,

$$\pi_{a|x} = \sum_{\tilde{\lambda}} D_{\tilde{\lambda}}(a|x)\, \sigma'_{a,\tilde{\lambda}}, \tag{E2}$$

for $\sigma'_{a,\tilde{\lambda}} \geq 0$, with $\text{Tr}[\sigma'_{a,\tilde{\lambda}}] = \text{Tr}[\sigma'_{a',\tilde{\lambda}}]$, $\forall a, a' \in A$. Both $D_{\tilde{\lambda}}(a|x)$ and $\tilde{\sigma}_{a,\tilde{\lambda}}$ are also defined as in Eq. (A1). The problem of finding t^* is then expressed explicitly as the following SDP:

$$\text{Given} \quad \mathbf{\Sigma}_{A|X} = \{\sigma_{a|x}\}_{a,x},$$

$$\min_{\{\tilde{\chi}_{a,\lambda}, \chi_{a,\lambda} \in \mathcal{B}(\mathbb{H}_B)\}_{a,\lambda}} \quad t, \tag{E3a}$$

$$\text{s. t.} \quad \sigma_{a|x} = \sum_{\lambda} D_{\lambda}(a|x)\, (\tilde{\chi}_{a,\lambda} - \chi_{a,\lambda}), \tag{E3b}$$

$$\text{with} \quad 1 + t = \sum_{\lambda} \text{Tr}[\tilde{\chi}_{a,\lambda}], \tag{E3c}$$

$$\text{Tr}[\tilde{\chi}_{a,\lambda}] = \text{Tr}[\tilde{\chi}_{0,\lambda}], \tag{E3d}$$

$$\text{Tr}[\chi_{a,\lambda}] = \text{Tr}[\chi_{0,\lambda}], \tag{E3e}$$

$$\tilde{\chi}_{a,\lambda} \geq 0, \tag{E3f}$$

$$\text{and} \quad \chi_{a,\lambda} \geq 0. \tag{E3g}$$

Note that normalization of $\sigma_{a|x}$ automatically implies $\sum_{\lambda} \text{Tr}[\chi_{a,\lambda}] = t$, which explains why one does not impose it as an independent, explicit constraint on the optimization.

Like Eq. (A3), the test (E3) admits a dual formulation, which takes the following form.

$$\text{Given} \quad \mathbf{\Sigma}_{A|X} = \{\sigma_{a|x}\}_{a,x},$$

$$\max_{\{V_{a|x} \in \mathcal{B}(\mathbb{H}_B)\}_{a,x}} \quad \sum_{a,x} \text{Tr}[V_{a|x}\, \sigma_{a|x}] - 1, \tag{E4a}$$

$$\text{s. t.} \quad B'_{a,\lambda}\, \mathbb{1}_B \leq \sum_{x} V_{a|x}\, D_{\lambda}(a|x) \leq B_{a,\lambda}\, \mathbb{1}_B, \tag{E4b}$$

$$\text{with} \quad \sum_{a} B_{a,\lambda} = 1, \tag{E4c}$$

$$\text{and} \quad \sum_{a} B'_{a,\lambda} = 0. \tag{E4d}$$

where $B_{a,\lambda}$, $B'_{a,\lambda} \in \mathbb{R}$ for all $a \in \mathbb{Z}_A$ and all $\lambda \in \mathbb{Z}_\Lambda$. With the same arguments as in the discussion right after Eqs. (A4), one sees that Eq. (E4a) returns a positive maximum (the one defining t^*) if, and only if, $\mathbf{\Sigma}_{A|X} \notin$ 1SQI. In fact, using that $\mathbf{\Sigma}_{A|X}$ is well-normalised, the term -1 in the objective function can be absorbed into the witness' definition with the variable change $\mathbf{V}_{A|X} \to \mathbf{W}_{A|X} := \{V_{a|x} - \mathbb{1}_B/|X|\}_{a,x}$. The resulting SDP (for the redefined witness $\mathbf{W}_{A|X}$) is similar to the one in Eqs. (A4), but with an extra constraint coming from the left-hand side inequality of Eq. (E4b), and with the witness scale no longer fixed. Thus, the robustness is given by the violation of the optimal $\mathbf{V}_{A|X}$ over all non-instrumentality witnesses with 1SQI bound $\beta = 1$ and subject to the specific constraints given by Eqs. (E4).

Appendix F: Proof of theorem 2

Consider an arbitrary $\mathbf{\Sigma}_{A|X}^{(\text{qinst})} \in$ QI. By definition, it admits a decomposition as in Eq. (2). Here, we use the short-hand notation $\mathbf{\Sigma}_{A|X}^{(\text{qinst})} = \mathcal{E}_{B|A,\Lambda_B}(\mathbf{\Sigma}_{A|X}^{(\text{q})})$, where $\mathcal{E}_{B|A,\Lambda_B} := \{\mathcal{E}_{B|a,\Lambda_B}\}_a$, to represent Eq. (2). In addition, we denote by $\mathcal{E}_{B|A,\Lambda_B}^* := \{\mathcal{E}_{B|a,\Lambda_B}^*\}_a$ the collection of dual (adjoint) maps $\mathcal{E}_{B|a,\Lambda_B}^*$ of each completely-positive trace-preserving (CPTP) map $\mathcal{E}_{B|a,\Lambda_B}$. This has the property that $\langle \mathbf{V}_{A|X}, \mathcal{E}_{B|A,\Lambda_B}(\mathbf{\Sigma}_{A|X}) \rangle = \langle \mathcal{E}_{B|A,\Lambda_B}^*(\mathbf{V}_{A|X}), \mathbf{\Sigma}_{A|X} \rangle$, for any $\mathbf{V}_{A|X}$ and $\mathbf{\Sigma}_{A|X}$. Then, if $\mathbf{V}_{A|X}$ is the optimal robustness witness of $\mathbf{\Sigma}_{A|X}^{(\text{qinst})}$, defined by Eqs. (E4), it holds

that

$$R_{\text{ni}}\left(\mathbf{\Sigma}_{A|X}^{(\text{qinst})}\right) = \left\langle \mathbf{V}_{A|X}, \mathbf{\Sigma}_{A|X}^{(\text{qinst})}\right\rangle - 1$$

$$= \left\langle \mathbf{V}_{A|X}, \mathcal{E}_{B|A,\Lambda_B}\left(\mathbf{\Sigma}_{A|X}^{(\text{q})}\right)\right\rangle - 1$$

$$= \left\langle \mathcal{E}_{B|A,\Lambda_B}^{*}\left(\mathbf{V}_{A|X}\right), \mathbf{\Sigma}_{A|X}^{(\text{q})}\right\rangle - 1. \quad \text{(F1)}$$

Now, assume, for a moment, that $\mathcal{E}_{B|A,\Lambda_B}^{*}\left(\mathbf{V}_{A|X}\right)$ is also a valid robustness witness. Then, denoting by $\mathbf{U}_{A|X}$ the optimal robustness witness for $\mathbf{\Sigma}_{A|X}^{(\text{q})}$, it must hold that

$$\left\langle \mathcal{E}_{B|A,\Lambda_B}^{*}\left(\mathbf{V}_{A|X}\right), \mathbf{\Sigma}_{A|X}^{(\text{q})}\right\rangle - 1 \leq \left\langle \mathbf{U}_{A|X}, \mathbf{\Sigma}_{A|X}^{(\text{q})}\right\rangle - 1.$$
$$\text{(F2)}$$

The left-hand side of this equation equals $R_{\text{ni}}\left(\mathbf{\Sigma}_{A|X}^{(\text{qinst})}\right)$, whereas the right-hand side equals $R_{\text{ni}}\left(\mathbf{\Sigma}_{A|X}^{(\text{q})}\right)$, thus giving $R_{\text{ni}}\left(\mathbf{\Sigma}_{A|X}^{(\text{q})}\right) \geq R_{\text{ni}}\left(\mathbf{\Sigma}_{A|X}^{(\text{qinst})}\right)$. So, the only missing ingredient is to show that $\mathcal{E}_{B|A,\Lambda_B}^{*}\left(\mathbf{V}_{A|X}\right)$ is, in fact, a valid robustness witness.

To prove this we note that, for any $\mathbf{\Sigma}_{A|X}^{(\text{inst})} \in$ 1SQI, $\mathcal{E}_{B|A,\Lambda_B}(\mathbf{\Sigma}_{A|X}^{(\text{inst})})$ is also in 1SQI. Hence, $\left\langle \mathcal{E}_{B|A,\Lambda_B}^{*}\left(\mathbf{V}_{A|X}\right), \mathbf{\Sigma}_{A|X}^{(\text{inst})}\right\rangle = \left\langle \mathbf{V}_{A|X}, \mathcal{E}_{B|A,\Lambda_B}\left(\mathbf{\Sigma}_{A|X}^{(\text{inst})}\right)\right\rangle \leq 1$, for all $\mathbf{\Sigma}_{A|X}^{(\text{inst})} \in$ 1SQI. This implies that $\mathcal{E}_{B|A,\Lambda_B}^{*}\left(\mathbf{V}_{A|X}\right)$ is a non-instrumentality witness with $\beta_{\mathcal{E}_{B|A,\Lambda_B}^{*}(\mathbf{V}_{A|X})} = 1$. Also, given that each $\mathcal{E}_{B|a,\Lambda_B}^{*}$ is completely-positive (CP) and unital, since it is the dual of a CPTP map, applying these dual maps to any robustness witness does not invalidate its defining constraints, in Eqs. (E4b)-(E4d). This means that the resulting object after the application of the dual map is also a valid 1SQI-robustness witness in the SDP formulation. $\quad\square$