

# Neither Contextuality nor Nonlocality Admits Catalysts

Martti Karvonen  
*University of Ottawa, Canada*

We show that the resource theory of contextuality does not admit catalysts, i.e., there are no correlations that can enable an otherwise impossible resource conversion and still be recovered afterward. As a corollary, we observe that the same holds for nonlocality. As entanglement allows for catalysts, this adds a further example to the list of “anomalies of entanglement,” showing that nonlocality and entanglement behave differently as resources. We also show that catalysis remains impossible even if, instead of classical randomness, we allow some more powerful behaviors to be used freely in the free transformations of the resource theory.

*Introduction.*—Contextuality [1, 2] and nonlocality [3, 4] play a prominent role in a wide variety of applications of quantum mechanics, with nonlocality being used, for example, in quantum key-distribution [5], certified randomness [6] and randomness expansion [7]. Similarly, contextuality powers quantum computation in some computational models [8–13] and even increases expressive power in quantum machine learning [14]. Consequently, it is vital to understand how nonlocality and contextuality behave as resources.

In this Letter, we show that neither contextuality nor nonlocality admits catalysts: that is, there are no correlations that can be used to enable an otherwise impossible conversion between correlations and still be recovered afterward. Slightly more precisely, let us write  $d, e, f \dots$  for various correlations (whether classical or not),  $d \otimes e$  for having independent instances of  $d$  and  $e$ , and  $d \rightsquigarrow e$  (read as “ $d$  simulates  $e$ ”) for the existence of a conversion  $d \rightarrow e$ . Then our results state that, in suitably formalized resource theories of contextuality and nonlocality, whenever  $d \otimes e \rightsquigarrow d \otimes f$ , then  $e \rightsquigarrow f$  already. This gives a strong indication that contextuality (and nonlocality) are resources that *get spent* when you use them: there is no way of using a correlation  $d$  to achieve a task you could not do otherwise while keeping  $d$  intact. As entanglement theory famously allows for catalysts [15], this can be seen as yet another “anomaly of nonlocality” [16] and thus further testament to the fact that nonlocality and entanglement are different resources.

We prove our results by working in precisely defined resource theories of contextuality and nonlocality. These are not strictly speaking quantum resource theories [17], but resource theories in a more general sense [18, 19], as we allow resources such as Popescu-Rohrlich boxes (PR boxes) [20] that are not quantum realizable. The kinds of conversions between correlations we have in mind capture the intuitive idea of using one system to *simulate* another one, and have been studied in earlier literature [21–29]. These roughly correspond to the local operations and shared randomness paradigm (LOSR) or to wirings and prior-to-input classical communication, depending on the precise definitions of these terms. However, existing formalizations of these in the literature are often limited

to the bipartite or tripartite settings, and at times overlook some technical issues resulting in nonconvex sets of transformations [30, Appendix].

More importantly, existing formalizations of the resource theory of nonlocality tend to focus on the case where each party has a discrete set of measurements of which they can perform at most one. However, this is false even in relatively simple situations: for instance, if Alice shares one PR box with Bob and one with Charlie, then she has four measurements available but is not restricted to only one measurement as she can choose a measurement for each of her boxes. In particular, she might first measure one of the boxes and use the outcome (and possible auxiliary randomness) to choose what to measure next.

To overcome such issues, we work in the general approach to contextuality initiated in [31] and later extended in [32] to capture building some correlations from others using such *probabilistic* and *adaptive* means, resulting in a resource theory for contextuality. We then obtain the resource theory of nonlocality from this via a general mathematical construction used in [33] that builds a resource theory of  $n$ -partite resources from a given resource theory. Working with such generality clarifies the relationship between resource theories of contextuality and nonlocality and captures the kinds of interconversions studied in earlier literature in the exact, single-shot regime [34] in a precise yet tractable manner.

We believe that our result could be phrased and proved in terms other approaches to contextuality [35–39], as long as one formalizes such adaptive measurement protocols and transformations between correlations within them. However, the current proof strategy no longer applies if one works with axiomatically defined transformations as in [40], i.e. with abstract functions between sets of correlations satisfying properties such as preservation of locality and of convex combinations. This is because such abstract functions might not arise from operationally defined protocols that one might implement physically. Indeed, operational transformations form a proper subset of axiomatically defined ones for the resource theories of entanglement [41] and magic [42]. For contextuality and nonlocality it is not known if the ax-

omatic and operational resource theories agree, although a characterization of those functions arising from (non-adaptive) operational transformations is given in [43, Theorem 44].

One possible explanation for anomalies of nonlocality, put forward in [44], is that they stem from using LOSR transformations with nonlocality and local operations and classical communication (LOCC) with entanglement. Indeed, [44] shows that many of the anomalies disappear when working with LOSR entanglement. In particular they show, for bipartite pure states, that there are no catalysts for LOSR entanglement. We conjecture that there are no catalysts in general for LOSR entanglement, in which case this anomaly is fully explained by entanglement and nonlocality being measured with LOCC and LOSR transformations respectively. For a contrasting viewpoint, see [45] which argues that both entanglement and nonlocality should be measured in terms of LOCC transformations.

*No catalysis for contextuality.*—We begin by briefly reviewing the resource theory of contextuality as defined in [32]. To start, we formalize the idea of a “measurement scenario”  $S$ : we imagine a situation where there is a finite set  $X_S$  of measurements available, each measurement  $x \in X_S$  giving rise to outcomes in some finite set  $O_{S,x}$ . However, only some measurements might be feasible to perform together—other combinations may be ruled out by practical limitations or excluded by physical theory. We collect all jointly compatible measurements into a single set  $\Sigma_S$ , which we expect to satisfy two natural properties: (1) any measurement  $x \in X_S$  induces a compatible set  $\{x\} \in \Sigma_S$ , and (2) any subset of a compatible set of measurements is compatible. Collecting all this data together results in the measurement scenario  $S = \langle X_S, \Sigma_S, O_S \rangle$ .

Given two scenarios  $S$  and  $T$ , we let  $S \otimes T$  denote the scenario that represents having access to  $S$  and  $T$  in parallel, so that a joint measurement is possible precisely if its components in  $S$  and  $T$  are possible.

An *empirical model* over a scenario  $S$  is given by specifying for each compatible  $\sigma \in \Sigma_S$  a joint probability distribution  $e_\sigma$  for measurements in  $\sigma$ . We only consider empirical models for which the behavior of a joint measurement does not depend on what is measured with it, if anything. Thus, whenever  $\tau \subset \sigma \in \Sigma_S$ , the distribution  $e_\tau$  over  $\tau$  can be obtained by marginalizing  $e_\sigma$  to  $\tau$ . We express this generalization of the usual no-signaling conditions as  $e_\sigma|_\tau = e_\tau$ , so that more generally for a joint distribution  $d$  over outcomes of  $Y \subset X_S$  and  $Z \subset Y$ , the expression  $d|_Z$  denotes the marginal distribution on outcomes of  $Z$ .

If the impossibility of measuring everything together is only a practical limitation, one can contemplate the distribution  $d$  that would arise when measuring  $X_S$ . If  $d$  explained the model  $e$  we have at hand, we would expect it to satisfy  $d|_\sigma = e_\sigma$  for every  $\sigma \in \Sigma_S$ . If such a

distribution exists, we call  $e$  “noncontextual”. If no such distribution exists, we call  $e$  “contextual”, as we have reason to believe that it is infeasible in principle to measure everything together, unless one accepts that the observed joint distribution for a subset  $Y \subset X$  depends on what it is measured with i.e., its context.

A simple example of scenario and a contextual empirical model on it, discussed in [46], is given by three measurements, any two of which are compatible but not all three together. Each measurement takes outcomes in  $\{0, 1\}$ , and whenever two measurements are performed one observes  $(0, 1)$  and  $(1, 0)$  with equal probability, with the probability distributions for singletons fixed by this and resulting in observing the two outcomes with equal probability. To see that this is contextual, note that there is no joint outcome for all three measurements that is consistent with the observed marginals.

More examples can be obtained from scenarios studied in nonlocality, where one typically specifies a scenario by giving the number of parties, the number of measurements available to each of them and the size of the outcome sets, where it is then understood that maximal compatible measurements are given by a choice of a single measurement by each party. This includes for instance the famous Clauser, Home, Shimony, and Holt model [47] and the PR box [20] which goes beyond what is allowed in quantum mechanics, both models arising in the scenario with two parties having access to two dichotomic measurements.

We now move on to transformations between scenarios and empirical models. We will build up to “wirings” that in full generality capture the idea of “simulating” an empirical model from another adaptively with the help of noncontextual randomness. These will be the free transformations of our resource theory, but we begin by considering the problem of building one scenario from another. A particularly simple way of building  $T$  from  $S$ , is by declaring that for each measurement  $x \in X_T$  of  $T$  some measurement  $\pi(x) \in X_S$  of  $S$  is to be performed instead. Moreover, each outcome  $o$  of  $\pi(x)$  is to be interpreted as the outcome  $\alpha_x(o)$  instead. If for each compatible  $\sigma \in \Sigma_S$  the corresponding measurement  $\pi(\sigma)$  is jointly compatible in  $S$ , the pair  $\langle \pi, \alpha = (\alpha_x)_{x \in X} \rangle$  describes a way of building  $T$  from  $S$ , and we will denote this by writing  $\langle \pi, \alpha \rangle : S \rightarrow T$ .

Given such  $\langle \pi, \alpha \rangle : S \rightarrow T$ , any empirical model  $e : S$  induces a model on  $T$  that describes the statistics one would see if one was to observe the statistics given by  $e$  and then transform them according to  $\langle \pi, \alpha \rangle$ . We then denote by  $\langle \pi, \alpha \rangle_*$  the function that pushes empirical models on  $S$  forward to empirical models on  $T$ , so that for  $e : S$  the induced model on  $T$  is denoted by  $\langle \pi, \alpha \rangle_*(e)$ . The pair  $\langle \pi, \alpha \rangle$  is defined to be a “deterministic simulation” of  $e : T$  from  $d : S$ , denoted by  $\langle \pi, \alpha \rangle : d \rightarrow e$ , precisely when  $\langle \pi, \alpha \rangle$  transforms  $d$  to  $e$ , i.e., if  $\langle \pi, \alpha \rangle_*(d) = e$ . Such a simulation is depicted in Fig. 1

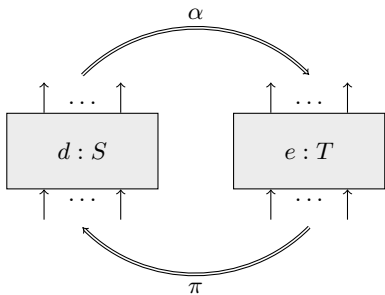


FIG. 1. Depiction of a deterministic simulation  $d \rightarrow e$ , with  $\pi$  transforming inputs of  $T$  to those of  $S$  and  $\alpha$  transforming outputs of  $S$  to outputs of  $T$ .

There are two ways in which these deterministic simulations are weaker than one would want in a general resource theory of contextuality: first, one might want to allow the usage of auxiliary (noncontextual) randomness, so that the dependence of measurements and their outcomes of  $T$  on those of  $S$  is stochastic. The usage of auxiliary randomness is captured in [48] and discussed further in [49], and in our terms, could be defined by allowing probabilistic mixtures of transformations. However, this viewpoint leaves out another important generalization: namely the possibility that a single measurement in  $T$  can depend on a joint measurement of  $S$  as in [50] or more generally on a measurement protocol on  $S$  that chooses which joint measurement to perform *adaptively*. The most general formulation of this idea would allow a measurement  $x$  in  $T$  to be simulated by a probabilistic and *adaptive* procedure, that first measures (depending on some classical randomness) something in  $S$  and then, based on the outcome (and possibly further classical randomness), chooses what to measure next (if anything) and so on. This idea is formalized carefully in [32] in two stages, the first one adding adaptivity and the second adding randomness.

To model adaptivity, one builds from a scenario  $S$  a new scenario  $\text{MP}(S)$  of (deterministic) “measurement protocols” over  $S$ . A measurement protocol is a procedure that, at any stage, either stops and reports all of the measurement results obtained so far, or, based on previously seen outcomes, performs a measurement in  $S$  that is compatible with the previous measurements. The measurements of  $\text{MP}(S)$  are given by such protocols over  $S$ , and a set of measurement protocols is compatible if they can be performed jointly without having to query measurements outside of  $\Sigma_S$ . Then one can define adaptive (but still deterministic) transformations  $S \rightarrow T$  between scenarios as deterministic transformations  $\text{MP}(S) \rightarrow T$  as in Fig. 2. In [32], we show that the assignment  $S \mapsto \text{MP}(S)$  defines a comonad on the category of scenarios: this abstract language is not needed here but can be thought of as guaranteeing that one has a well-behaved way of composing  $\text{MP}(S) \rightarrow T$  with

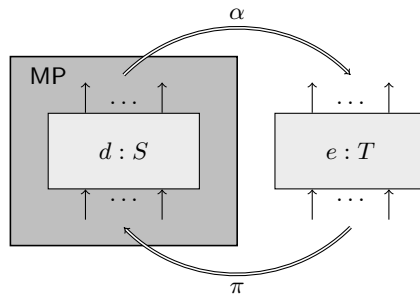


FIG. 2. Depiction of an adaptive simulation  $d \rightarrow e$

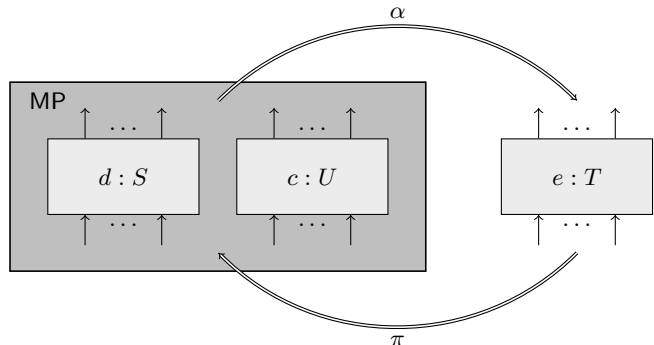


FIG. 3. General simulation  $d \rightarrow e$ , where we require  $c$  to be noncontextual.

$\text{MP}(T) \rightarrow U$  to obtain a map  $\text{MP}(S) \rightarrow U$ . Intuitively, the composite is obtained by first interpreting each measurement in  $U$  as a measurement protocol over  $T$ , and each measurement in that measurement protocol as an measurement protocol over  $S$ , and then “flattening” the resulting measurement protocol of measurement protocols {i.e., a measurement of  $\text{MP}[\text{MP}((S))]$ } into a measurement protocol over  $S$ .

To model (noncontextual) randomness, one then defines a simulation  $d \rightarrow e$  to be a deterministic simulation  $\text{MP}(d \otimes c) \rightarrow e$  for some noncontextual model  $c$  as in Fig. 3. Again, such simulations compose in a well-behaved manner, i.e. they form a category [51].

We will denote the existence of such a simulation  $d \rightarrow e$  by  $d \rightsquigarrow e$ , read as “ $d$  simulates  $e$ .” Simulations thus defined interact well with contextuality. For instance, [32, Theorem 21] states that the noncontextual fraction, studied in [13], is a monotone, that is, if  $d \rightsquigarrow e$ , then  $\text{NCF}(d) \leq \text{NCF}(e)$ ; and [50, Theorem 4.1] implies that an empirical model is noncontextual if and only if it can be simulated from the trivial model on the empty scenario. In fact, the notions of logical and strong contextuality can be captured along similar lines [43] by relaxing the equality of probability distributions in the definition of simulation by equality (or inclusion) of supports of these distributions.

We now state our main result for this resource theory of contextuality.

**Theorem 1.** *If  $d \otimes e \rightsquigarrow d \otimes f$  in the resource theory of contextuality, then  $e \rightsquigarrow f$ .*

The key ideas of the proof are simple, even if the full details get technical: if  $d$  can catalyze a transformation  $e \rightarrow f$  once, it can do so arbitrarily many times. Choosing a big enough number of copies of  $e$  to transform to copies of  $f$ , using the pigeonhole principle we can show that one only needs a compatible subset of  $S_d$  (or rather, a compatible set of measurement protocols). Making this precise requires formalizing our framework more carefully, and we do this in the Supplemental Material [52]. Our result subsumes [32, Theorem 22], as setting  $d = f$  and letting  $e$  be the trivial model on the empty scenario implies that if  $d \rightsquigarrow d \otimes d$ , the model  $d$  must be noncontextual.

*No-catalysis for nonlocality.*—We now explain how to interpret nonlocality within this framework. At the level of scenarios and models on them, nonlocality can be seen as a special case of contextuality: for nonlocality, the measurement scenario typically arises by considering  $n$  parties, with the  $i$ th party choosing one measurement from a set  $X_i$  of measurements available to them, with a measurement  $x \in X_i$  giving outcomes in some outcome set  $O_{i,x}$ . Often one restricts the situation even further and assumes that each party has the same number of measurements available and each measurement takes outcomes in a set of the same size, so that the scenario is specified (up to isomorphism) by three numbers: the number of parties, the number of measurements available to each of them and the size of the outcome sets. Whether or not one imposes this further restriction, such scenarios are of the form  $S = \bigotimes_{i=1}^n S_i$  where each  $S_i$  is just a discrete set of measurements (so only singleton measurements are possible in each  $S_i$ ). In particular, a maximal measurement corresponds to a choice  $(x_1, \dots, x_n)$  of a measurement at each site, and a set of correlations can be given as a family  $p(o_1, \dots, o_n | x_1, \dots, x_n)$  of conditional probabilities for each such measurement. If the family  $p$  is (fully) no-signaling, it corresponds to a unique empirical model  $e : S$  (where the probabilities over nonmaximal measurements are obtained by marginalization), and  $p$  is local if and only if  $e$  is noncontextual.

However, if we allow parties to share different non-local resources, we move away from the situation where each party chooses one measurement from a set of mutually exclusive measurements. For instance, if Bob shares one box with Alice and one with Charlie, Bob is not limited to a single measurement: he can choose one for each box. As quantum theory allows for arbitrary joint measurability graphs in the case of projection-valued measurements [53] and arbitrary simplicial complexes in the case of positive-operator valued measurements [54], we impose no restrictions on the measurement scenarios available at each site. However, if we are thinking about a  $n$ -partite scenario, it is reasonable to expect that measurement choices done

at one site do not affect the measurements available at another one. Thus, we model an  $n$ -partite scenario as a tuple  $(S_i)_{i=1}^n$  of scenarios, thought of as representing the  $n$  parties sharing the scenario  $\bigotimes_{i=1}^n S_i$ .

At the level of transformations between models and scenarios, nonlocality is no longer a special case of contextuality, as observed for example in [30, Appendix A.1]. This is because our wirings are slightly too general as they allow the  $i$ th party to wire some of their measurements to measurements belonging to other parties, whereas operationally speaking it is reasonable to require each party to have access only to measurements available to them (and shared randomness). Nevertheless, the resource theories are very closely related, as one can obtain the resource theory for  $n$ -partite nonlocality from that of contextuality via a general construction, used in the context of cryptography in [33], that builds a resource theory of  $n$ -partite resources from a given resource theory.

In our setting, this amounts to defining the resource theory of  $n$ -partite nonlocality as follows: an  $n$ -partite scenario is an  $n$ -tuple  $(S_i)_{i=1}^n$  of scenarios (some of which may be empty), and an  $n$ -partite empirical model  $e : (S_i)_{i=1}^n$  is an empirical model on  $\bigotimes_{i=1}^n S_i$ . The parallel composite  $\boxtimes$  of scenarios is defined pointwise, i.e., by setting  $(S_i)_{i=1}^n \boxtimes (T_i)_{i=1}^n = (S_i \otimes T_i)_{i=1}^n$ . To define  $\boxtimes$  for two  $n$ -partite models  $e : (S_i)_{i=1}^n$  and  $d : (T_i)_{i=1}^n$ , note that the scenarios  $(\bigotimes_{i=1}^n S_i) \otimes (\bigotimes_{i=1}^n T_i)$  and  $\bigotimes_{i=1}^n (S_i \otimes T_i)$  are canonically isomorphic so that the model  $e \otimes d : (\bigotimes_{i=1}^n S_i) \otimes (\bigotimes_{i=1}^n T_i)$  induces a model  $e \boxtimes d$  on  $\bigotimes_{i=1}^n (S_i \otimes T_i)$  via this isomorphism. Finally, an  $n$ -partite simulation  $d \rightarrow e$  is defined as an  $n$ -tuple of simulations  $[\text{MP}(T_i \otimes P_i \rightarrow S_i)]$  that, when taken together, transform  $d \boxtimes c$  to  $e$  where  $c$  is some noncontextual shared correlation. In this manner, the resource theory of  $n$ -partite nonlocality is derived from that of contextuality by keeping track of the  $n$ -partite nature of scenarios and transformations between them.

Our proof of Theorem 1 readily implies that nonlocality admits no catalysts.

**Theorem 2.** *If  $d \boxtimes e \rightsquigarrow d \boxtimes f$  in the resource theory of  $n$ -partite nonlocality, then  $e \rightsquigarrow f$ .*

*Quantum-assisted transformations and beyond.*—We briefly discuss a further generalization of our main result. One could consider resource theories with even more expressive simulations than the ones defined above. In the above, we define simulations as deterministic simulations assisted by noncontextual randomness. As suggested in [32], one could allow more general correlations to be used in transformations and study, e.g., quantum-assisted simulations. More specifically, one could define quantum-assisted simulations  $e \rightarrow f$  as deterministic simulations  $\text{MP}(e \otimes q) \rightarrow f$  where  $q$  is a quantum-realizable empirical model. More generally, for any class  $\mathcal{X}$  of empirical models that is closed under  $\otimes$ , one gets a well-defined notion of  $\mathcal{X}$ -assisted simulations, and we

will write  $d \rightsquigarrow_{\mathcal{X}} e$  to denote the existence of an  $\mathcal{X}$ -assisted simulation from  $d$  to  $e$ . If the class  $\mathcal{X}$  contains all non-local correlations, the resulting resource theory has no catalysts. A similar result holds for  $\mathcal{X}$ -assisted  $n$ -partite transformations between  $n$ -partite correlations, but we restrict ourselves to stating the theorem for contextuality.

**Theorem 3.** *Let  $\mathcal{X}$  be a class of empirical models that is closed under  $\otimes$  and contains all classical correlations. Then  $d \otimes e \rightsquigarrow_{\mathcal{X}} d \otimes f$  implies  $e \rightsquigarrow_{\mathcal{X}} f$ .*

For instance, if we take  $\mathcal{X}$  to be the class of quantum-realizable empirical models, this result implies that, if by some miracle we got access to a single PR box, we could not use it to catalyze a quantum-assisted transformation that was hitherto impossible. Put another way, no matter what class of correlations we can use freely, the only way to get mileage out of a box that goes beyond our powers is to spend it—so we must choose wisely.

We wish to thank Marcelo Terra Cunha for asking whether contextuality admits catalysts and for Ehtibar Dzhafarov for organizing the Quantum Contextuality in Quantum Mechanics and Beyond workshop where we met Marcelo and became interested in this question. We also thank Samson Abramsky, Rui Soares Barbosa and Shane Mansfield for helpful discussions.

- 
- [1] S. Kochen and E. P. Specker, The Problem of Hidden Variables in Quantum Mechanics, *J. Math. Mech.* **17**, 59 (1967).
- [2] C. Budroni, A. Cabello, O. Gühne, M. Kleinmann, and J.-Å. Larsson, Quantum contextuality, arXiv:2102.13036.
- [3] J. S. Bell, On the problem of hidden variables in quantum mechanics, *Rev. Mod. Phys.* **38**, 447 (1966).
- [4] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014).
- [5] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev, The security of practical quantum key distribution, *Rev. Mod. Phys.* **81**, 1301 (2009).
- [6] A. Acín and L. Masanes, Certified randomness in quantum physics, *Nature* **540**, 213 (2016).
- [7] S. Fehr, R. Gelles, and C. Schaffner, Security and composability of randomness expansion from Bell inequalities, *Phys. Rev. A* **87**, 012335 (2013).
- [8] J. Anders and D. E. Browne, Computational Power of Correlations, *Phys. Rev. Lett.* **102**, 050502 (2009).
- [9] R. Raussendorf, Contextuality in measurement-based quantum computation, *Phys. Rev. A* **88**, 022322 (2013).
- [10] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the ‘magic’ for quantum computation, *Nature* **510**, 351 (2014).
- [11] J. Bermejo-Vega, N. Delfosse, D. E. Browne, C. Okay, and R. Raussendorf, Contextuality as a resource for Models of Quantum Computation with Qubits, *Phys. Rev. Lett.* **119**, 120505 (2017).
- [12] R. Raussendorf, D. E. Browne, N. Delfosse, C. Okay, and J. Bermejo-Vega, Contextuality and Wigner-function negativity in qubit quantum computation, *Phys. Rev. A* **95**, 052334 (2017).
- [13] S. Abramsky, R. S. Barbosa, and S. Mansfield, Contextual Fraction as a Measure of Contextuality, *Phys. Rev. Lett.* **119**, 050504 (2017).
- [14] X. Gao, E. R. Anschuetz, S.-T. Wang, J. I. Cirac, and M. D. Lukin, Enhancing generative models via quantum correlations, arXiv:2101.08354.
- [15] D. Jonathan and M. B. Plenio, Entanglement-Assisted Local Manipulation of Pure Quantum States, *Phys. Rev. Lett.* **83**, 3566 (1999).
- [16] A. A. Méthot and V. Scarani, An anomaly of non-locality, *Quantum Inf. Comput.* **7**, 157 (2007).
- [17] E. Chitambar and G. Gour, Quantum resource theories, *Rev. Mod. Phys.* **91**, 025001 (2019).
- [18] B. Coecke, T. Fritz, and R. W. Spekkens, A mathematical theory of resources, *Inf. Comp.* **250**, 59 (2016).
- [19] T. Fritz, Resource convertibility and ordered commutative monoids, *Math. Struct. Comput. Sci.* **27**, 850 (2017).
- [20] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Found. Phys.* **24**, 379 (1994).
- [21] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Nonlocal correlations as an information-theoretic resource, *Phys. Rev. A* **71**, 022101 (2005).
- [22] J. Barrett and S. Pironio, Popescu-Rohrlich Correlations as a Unit of Nonlocality, *Phys. Rev. Lett.* **95**, 140401 (2005).
- [23] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vértesi, Closed sets of nonlocal correlations, *Phys. Rev. A* **80**, 062107 (2009).
- [24] N. S. Jones and L. Masanes, Interconversion of nonlocal correlations, *Phys. Rev. A* **72**, 052312 (2005).
- [25] F. Dupuis, N. Gisin, A. Hasidim, A. A. Méthot, and H. Pilpel, No nonlocal box is universal, *J. Math. Phys.* **48**, 082107 (2007).
- [26] J. I. De Vicente, On nonlocality as a resource theory and nonlocality measures, *J. Phys. A* **47**, 424017 (2014).
- [27] M. Forster and S. Wolf, Bipartite units of nonlocality, *Phys. Rev. A* **84**, 042112 (2011).
- [28] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, Operational Framework for Nonlocality, *Phys. Rev. Lett.* **109**, 070401 (2012).
- [29] R. Gallego and L. Aolita, Nonlocality free wirings and the distinguishability between Bell boxes, *Phys. Rev. A* **95**, 032118 (2017).
- [30] E. Wolfe, D. Schmid, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Quantifying Bell: the Resource Theory of Nonclassicality of Common-Cause Boxes, *Quantum* **4**, 280 (2020).
- [31] S. Abramsky and A. Brandenburger, The sheaf-theoretic structure of non-locality and contextuality, *New J. Phys.* **13**, 113036 (2011).
- [32] S. Abramsky, R. S. Barbosa, M. Karvonen, and S. Mansfield, A comonadic view of simulation and quantum resources, in *Proceedings of 34th Annual ACM/IEEE Symposium on Logic in Computer Science (LiCS 2019)* (IEEE, 2019) pp. 1–12.
- [33] A. Broadbent and M. Karvonen, Categorical composable cryptography, arXiv:2105.05949.
- [34] This is in contrast to asymptotic questions [17, V.B] or

- distillation tasks [55] where one tries to approximate a target resource using increasing numbers of copies of the starting resource. It is unclear if the notion of a catalyst makes sense in such settings, which is why we work with single-shot convertibility.
- [35] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, *Phys. Rev. A* **71**, 052108 (2005).
- [36] E. N. Dzhafarov and J. V. Kujala, Contextuality is about identity of random variables, *Phys. Scr.* **2014**, 014009 (2014).
- [37] A. Cabello, S. Severini, and A. Winter, Graph-Theoretic Approach to Quantum Correlations, *Phys. Rev. Lett.* **112**, 040401 (2014).
- [38] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, A combinatorial approach to nonlocality and contextuality, *Commun. Math. Phys.* **334**, 533 (2015).
- [39] S. Staton and S. Uijlen, Effect algebras, presheaves, nonlocality and contextuality, in *Proceedings of 42nd International Colloquium on Automata, Languages, and Programming (ICALP 2015)*, Lecture Notes in Computer Science, edited by M. M. Halldórsson, K. Iwama, N. Kobayashi, and B. Speckmann (Springer, 2015) pp. 401–413.
- [40] P. Joshi, A. Grudka, K. Horodecki, M. Horodecki, P. Horodecki, and R. Horodecki, No-broadcasting of non-signalling boxes via operations which transform local boxes into local ones, *Quantum Inf. Comput.* **13**, 0567 (2013).
- [41] C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters, Quantum nonlocality without entanglement, *Phys. Rev. A* **59**, 1070 (1999).
- [42] A. Heimendahl, M. Heinrich, and D. Gross, The axiomatic and the operational approaches to resource theories of magic do not coincide, arXiv:2011.11651.
- [43] R. S. Barbosa, M. Karvonen, and S. Mansfield, Closing Bell: Boxing black box simulations in the resource theory of contextuality, arXiv:2104.11241.
- [44] D. Schmid, T. C. Fraser, R. Kunjwal, A. B. Sainz, E. Wolfe, and R. W. Spekkens, Understanding the interplay of entanglement and nonlocality: Motivating and developing a new branch of entanglement theory, arXiv:2004.09194.
- [45] K. Sengupta, R. Zibakhsh, E. Chitambar, and G. Gour, Quantum Bell nonlocality is entanglement, arXiv:2012.06918.
- [46] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Specker’s parable of the overprotective seer: A road to contextuality, nonlocality and complementarity, *Phys. Rep.* **506**, 1 (2011).
- [47] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed Experiment to Test Local Hidden-Variable Theories, *Phys. Rev. Lett.* **23**, 880 (1969).
- [48] B. Amaral, A. Cabello, M. T. Cunha, and L. Aolita, Non-contextual Wirings, *Phys. Rev. Lett.* **120**, 130403 (2018).
- [49] B. Amaral, Resource theory of contextuality, *Philos. Trans. R. Soc. A* **377**, 20190010 (2019).
- [50] M. Karvonen, Categories of empirical models, in *Proceedings of 15th International Conference on Quantum Physics and Logic (QPL 2018)*, Electronic Proceedings in Theoretical Computer Science, Vol. 287, edited by P. Selinger and G. Chiribella (2019) pp. 239–252.
- [51] The abstract explanation is that the MP comonad is comonoidal as observed in [32, Theorem 17]. Roughly speaking this boils down to the properties of the transformations  $\text{MP}(S \otimes T) \rightarrow \text{MP}(S) \otimes \text{MP}(T)$  which interpret measurement protocols in  $S$  or in  $T$  as measurement protocols over  $S \otimes T$ .
- [52] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.127.160402> or at the end of this document for formal development of background material and full proofs.
- [53] C. Heunen, T. Fritz, and M. L. Reyes, Quantum theory realizes all joint measurability graphs, *Phys. Rev. A* **89**, 032121 (2014).
- [54] R. Kunjwal, C. Heunen, and T. Fritz, Quantum realization of arbitrary joint measurability structures, *Phys. Rev. A* **89**, 052126 (2014).
- [55] N. Brunner and P. Skrzypczyk, Nonlocality Distillation and Postquantum Theories with Trivial Communication Complexity, *Phys. Rev. Lett.* **102**, 160403 (2009).

# Neither Contextuality nor Nonlocality Admits Catalysts—Supplemental Material

Martti Karvonen  
University of Ottawa, Canada

In this Supplemental Material we provide proofs of our main theorems and some technical details concerning the background that is needed in order to formalize these.

## Background

We mostly follow the development of [32] with some changes in notation. For a scenario  $S$  and subset  $Y \subset X_S$  of measurements, we let  $\mathcal{E}_S(Y)$  (or  $\mathcal{E}(Y)$  when  $S$  is clear from context) denote the set of possible joint outcomes for the (not necessarily comensurable) set  $Y$ . Formally,  $\mathcal{E}_S(Y)$  is defined as the cartesian product  $\prod_{x \in Y} O_{S,x}$ .

We now define empirical models more carefully. Given a probability distribution  $d$  on  $\mathcal{E}_S(Z)$ , which we think of as a joint probability distribution for measurements in  $Z$ , and a subset  $Y \subset Z$ , we denote by  $d|_Y$  the marginalization of  $d$  to measurements in  $Y$ . Now, an empirical model  $e : S$  consists of a family  $(e_\sigma)_{\sigma \in \Sigma_S}$ , where each  $e_\sigma$  is a probability distribution over the set  $\mathcal{E}_S(\sigma)$ , and whenever  $\tau \subset \sigma \in \Sigma_S$  we have

$$e_\sigma|_\tau = e_\tau$$

The *parallel composite*  $S \otimes T$  of  $S$  and  $T$  is defined by  $X_{S \otimes T} = X_S \sqcup X_T$ ,  $\Sigma_{S \otimes T} = \{\sigma \sqcup \tau \mid \sigma \in \Sigma_S, \tau \in \Sigma_T\}$  and setting the outcome set at  $x \in X_S \sqcup X_T$  to be  $O_{S,x}$  if  $x \in X_S$  and  $O_{T,x}$  if  $x \in X_T$ . Here  $X \sqcup Y$  denotes the disjoint union of the sets  $X$  and  $Y$ , and can be defined e.g. as  $X \sqcup Y := X \times \{0\} \cup Y \times \{1\}$ .

A deterministic procedure  $S \rightarrow T$  consists of a

- A simplicial function  $\pi : X_T \rightarrow X_S$ , i.e., a function  $\pi$  satisfying  $\pi(\sigma) \in \Sigma_S$  for every  $\sigma \in \Sigma_T$
- A family  $\alpha = (\alpha_x : O_{S,\pi(x)} \rightarrow O_{T,x})_{x \in X}$  of functions between outcome sets.

A procedure  $\langle \pi, \alpha \rangle : S \rightarrow T$  induces a mapping between empirical models so that each  $e : S$  is pushed forward to  $\langle \pi, \alpha \rangle_* e : T$ . Mathematically, the empirical model  $\langle \pi, \alpha \rangle_* e$  is defined at measurement  $\sigma \in \Sigma_T$  by

$$(\langle \pi, \alpha \rangle_* e)_\sigma = \alpha_*(e_{\pi\sigma})$$

where the right hand side denotes the pushforward of the probability distribution  $e_{\pi\sigma}$  along the function  $\mathcal{E}_S(\pi(\sigma)) \rightarrow \mathcal{E}_T(\sigma)$  whose  $x$ th coordinate projection is given by  $\mathcal{E}_S(\pi(\sigma)) \rightarrow \mathcal{E}_S(\pi(x)) \xrightarrow{\alpha_x} \mathcal{E}_T(x)$ .

If one imagines  $S$  as corresponding to some particular experimental setup, and a model  $e : S$  as corresponding to empirically observed propensities of outcomes for a fixed state preparation, then one can interpret a procedure  $\langle \pi, \alpha \rangle : S \rightarrow T$  as a way of (deterministically) building the experimental setup  $T$  out of  $S$ : the map  $\pi$  tells what measurement in  $S$  each measurement in  $T$  corresponds to, and  $\alpha$  tells how to interpret outcomes in  $S$  as outcomes in  $T$ . Then the model  $\langle \pi, \alpha \rangle_* e$  describes the statistics one would see if one was to observe the statistics given by  $e$  but transform them according to  $\langle \pi, \alpha \rangle$ . In other words, the probability  $\langle \pi, \alpha \rangle_* e$  gives to some fixed outcome over the joint measurement  $\sigma \in \Sigma_T$  is the sum the probabilities  $e$  gives to all outcomes of  $\pi(\sigma)$  that are mapped to  $s$  by  $\alpha$ .

A deterministic simulation  $d \rightarrow e$  where  $d : S$  and  $e : T$  is a deterministic procedure  $\langle \pi, \alpha \rangle : S \rightarrow T$  that transforms  $d$  to  $e$ , i.e., that satisfies  $\langle \pi, \alpha \rangle_* d = e$ .

We now formalize measurement protocols carefully. These were used in [38, Appendix D] to relate the sheaf-theoretic approach [31] to contextuality to the hypergraph approach [38], and then later used in [32] to extend the former. Intuitively, a (deterministic) measurement protocol is a set of rules that tells at each stage, what to measure next given the previous measurements and their outcomes. At any stage of the measurement protocol, we thus have a sequence  $(x_i, o_i)_{i=1}^n$  of measurement-outcome pairs. A technical insight of [32] is that measurement protocols can be defined in terms of sets of such sequences: if you know all such sequences that could happen during a deterministic protocol, you also know the protocol itself.

**Definition 4.** A *run* on a measurement scenario  $S$  is a sequence  $\bar{x} := (x_i, o_i)_{i=1}^n$  such that  $x_i \in X_S$  are distinct,  $\{x_1, \dots, x_n\} \in \Sigma_S$ , and each  $o_i \in O_{S,x_i}$ . We denote the empty run by  $\Lambda$ .

A run  $\bar{x}$  determines a context  $\sigma_{\bar{x}} := \{x_1, \dots, x_n\} \in \Sigma_S$  and a joint assignment  $s_{\bar{x}} \in \mathcal{E}(\sigma_{\bar{x}})$  on that context that maps  $x_i$  to  $o_i$ . Two runs  $\bar{x}$  and  $\bar{y}$  are said to be *consistent* if they agree on common measurements, i.e., for every  $z \in \sigma_{\bar{x}} \cap \sigma_{\bar{y}}$  we have  $s_{\bar{x}}(z) = s_{\bar{y}}(z)$ .

Given runs  $\bar{x}$  and  $\bar{y}$ , we denote their concatenation by  $\bar{x} \cdot \bar{y}$ . Note that  $\bar{x} \cdot \bar{y}$  might not be a run.

**Definition 5.** A *measurement protocol* on  $S$  is a non-empty set  $Q$  of runs satisfying the following conditions:

- (i) if  $\bar{x} \cdot \bar{y} \in Q$  then  $\bar{x} \in Q$ ;

- (ii) if  $\bar{x} \cdot (x, o) \in Q$ , then  $\bar{x} \cdot (x, o') \in Q$  for every  $o' \in O_{S,x}$ ;
- (iii) if  $\bar{x} \cdot (x, o) \in Q$  and  $\bar{x} \cdot (x', o') \in Q$ , then  $x = x'$ .

Identifying a measurement protocol with its set of possible runs, the first condition guarantees that a prefix of a possible run is a possible run. The second condition ensures that, if  $x$  might be measured at some stage, then any outcome of  $x$  is an (in-principle) possible outcome at that stage, and the third condition guarantees that at any given stage, the next measurement prescribed by the protocol (if any), is uniquely determined.

In this viewpoint, we may identify an outcome of a measurement protocol  $Q$  with a *maximal run* of  $Q$ , i.e., a run  $\bar{x} \in Q$  that is not a proper prefix of any  $\bar{y} \in Q$ . Such a run describes a sequence of measurement-outcome pairs obtained during  $Q$ , and maximality of the run ensures that the measurement protocol was followed to its conclusion.

**Definition 6.** Given a scenario  $S$ , we build a scenario  $\text{MP}(S)$  as follows:

- its set of measurements is the set  $X_{\text{MP}(S)}$  of measurement protocols on  $S$ ;
- the outcome set  $O_Q$  of a measurement protocol  $Q \in X_{\text{MP}(S)}$  is its set of maximal runs, i.e., those  $\bar{x} \in Q$  that are not a proper prefix of any  $\bar{y} \in Q$ ;
- a set  $\{Q_1, \dots, Q_n\}$  of measurement protocols is compatible whenever for any choice of pairwise consistent runs  $\bar{x}_i \in Q_i$  with  $i \in \{1, \dots, n\}$ , we have  $\bigcup_i \sigma_{\bar{x}_i} \in \Sigma$ .

Given an empirical model  $e : S$ , we define the empirical model  $\text{MP}(e) : \text{MP}(S)$  as follows. For a compatible set  $\sigma := \{Q_1, \dots, Q_n\}$  of measurement protocols and an assignment  $s : Q_i \mapsto \bar{x}_i \in \mathcal{E}(\sigma)$ , we set

$$\text{MP}(e)_\sigma(s) := \begin{cases} e_{\bigcup_i \sigma_{\bar{x}_i}}(\bigcup_i s_{\bar{x}_i}) & \text{if } \{\bar{x}_i\} \text{ pairwise consistent} \\ 0 & \text{otherwise.} \end{cases}$$

The intuition behind a set of measurement protocols being compatible is that they can be performed together, without ever having to perform a measurement not allowed by  $\Sigma_S$ . We next make this intuition precise.

**Definition 7.** We say that a measurement protocol  $P$  contains implicitly a protocol  $Q$ , if any outcome of  $P$  determines the outcome of  $Q$ , and we denote this by  $P \succeq Q$ . Formally,  $P \succeq Q$  if for any maximal run  $\bar{x} \in P$  there is a (necessarily unique) maximal run  $\bar{y} \in Q$  such that the assignment  $s_{\bar{y}}$  determined by  $\bar{y}$  is a restriction of  $s_{\bar{x}}$ . We say that a measurement protocol  $P$  contains implicitly a set of protocols  $\{Q_1, \dots, Q_n\}$  if  $P \succeq Q_i$  for each  $i$ .

**Lemma 8.** *If  $\{Q_1, \dots, Q_n\}$  is a compatible set of measurement protocols, there is a measurement protocol  $P$  that implicitly contains them all.*

*Proof.* We define a protocol  $P$  that first performs  $Q_1$ , then (whatever is left of)  $Q_2$  and so on. Formally, we first define inductively a merge operation  $*$  for compatible runs:

$$\begin{aligned} \bar{x} * \Lambda &:= x \\ \bar{x} * ((y, o) \cdot \bar{y}) &:= \begin{cases} \bar{x} * \bar{y} & \text{if } y \in \sigma_{\bar{x}} \\ (\bar{x} \cdot (y, o)) * \bar{y} & \text{otherwise.} \end{cases} \end{aligned}$$

We extend  $*$  to all pairs of runs by setting  $\bar{x} * \bar{y} = \Lambda$  whenever  $\bar{x}$  and  $\bar{y}$  are not compatible.

We then define  $P$  by taking the closure of the set

$$\{\bar{x}_1 * \dots * \bar{x}_n \mid \bar{x}_i \in O_{\text{MP}(S), Q_i}\}$$

under prefixes. Now property (i) is true by construction, and properties (ii) and (iii) follow from each  $Q_i$  being a measurement protocol. Finally, each outcome of  $P$  is of the form  $\bar{x}_1 * \dots * \bar{x}_n$ , so that  $P \succeq Q_i$  for each  $Q_i$ .  $\square$

*Remark 9.* If  $P$  is a measurement protocol over  $S$  and we fix the outcomes of some subset  $Y \subset X_S$  of measurements to equal  $t \in \mathcal{E}(Y)$ , this determines a measurement protocol  $P(t)$  that does not measure anything in  $Y$  as follows: it proceeds exactly as  $P$  except that whenever  $P$  is supposed to measure some  $x \in Y$  it behaves as if  $t(x)$  was observed and proceeds accordingly. In particular, if  $P$  is a measurement protocol over  $S \otimes T$  and  $t$  is an element of  $\mathcal{E}_T(X_T)$ , then  $P(t)$  is a measurement protocol over  $S$ .

We can describe  $P(t)$  more formally as follows. Let us say that a run  $\bar{x} \in P$  is compatible with  $t \in \mathcal{E}(Y)$  if  $s_{\bar{x}}$  agrees with  $t$  on common measurements, i.e., if  $s_{\bar{x}}(z) = t(z)$  for every  $z \in \sigma_{\bar{x}} \cap Y$ . Given a run  $\bar{x}$  and  $t \in \mathcal{E}(Y)$ , we define  $\bar{x} \setminus t$  to be the run that omits measurements in  $Y$ . Formally, we can define this inductively on  $\bar{x}$  by setting

$$\begin{aligned} \Lambda \setminus t &= \Lambda \\ (\bar{x} \cdot (y, o)) \setminus t &:= \begin{cases} \bar{x} \setminus t & \text{if } y \in Y \\ (\bar{x} \setminus t) \cdot (y, o) & \text{otherwise.} \end{cases} \end{aligned}$$

Then we can define  $P(t)$  by

$$P(t) := \{\bar{x} \setminus t \mid \bar{x} \in P \text{ is compatible with } t\}$$

## Proofs

### Proof of Theorem 1

Consider models  $d : S$ ,  $e : T$  and  $f : U$  and an adaptive simulation  $d \otimes e \rightarrow d \otimes f$ . We wish to produce a simulation  $e \rightarrow f$ . Heuristically speaking, our overall strategy is to show that the composite  $d \otimes e \rightarrow d \otimes f \rightarrow f$  needs to use  $d$  only in a noncontextual manner. More precisely,



we show that a compatible set of measurement protocols over  $S$  suffices to carry out a simulation  $d \otimes e \rightarrow f$ . There are several compatible sets one could use for this: we choose one that works both for the resource theory of contextuality and for nonlocality.

We first reduce to the deterministic case: by definition, there is a deterministic (but adaptive) map  $d \otimes e \otimes c \rightarrow d \otimes f$  with  $c$  noncontextual. Let us set  $g := e \otimes c$  and denote the scenario of  $e \otimes c$  by  $R$ . We then have a deterministic map  $d \otimes g \rightarrow d \otimes f$ . Now, if  $d$  can catalyze the transformation  $g \rightarrow f$  once without getting spent in the process, it can do so arbitrarily many times by first catalyzing the first copy of  $g$  to  $f$ , then the second one and so on. This results in a deterministic simulation  $d \otimes g^{\otimes n} \rightarrow d \otimes f^{\otimes n}$  for any  $n$  with the property that simulating the  $i$ th copy of  $f$  uses only  $d$  and the  $i$ th copy of  $g$ .

As the composite  $d \otimes g^{\otimes n} \rightarrow d \otimes f^{\otimes n} \rightarrow f$  to the  $i$ th copy of  $f$  uses only one copy of  $g$  it hence induces a simulation  $\langle \pi_i, \alpha_i \rangle: d \otimes g \rightarrow f$ . As there are only finitely many deterministic procedures  $\text{MP}(S \otimes R) \rightarrow U$ , we can force as many of the simulations  $\langle \pi_i, \alpha_i \rangle$  to coincide by choosing large enough  $n$ . In particular, there is a single simulation that can work as many times as needed. More precisely, there is some fixed deterministic adaptive simulation  $\langle \pi, \alpha \rangle: d \otimes g \rightarrow f$ , such that for any  $n$  there is an deterministic adaptive simulation  $d \otimes g^{\otimes n} \rightarrow d \otimes f^{\otimes n}$  with the property that simulating the  $i$ th copy of  $f$  uses only  $d$  and the  $i$ th copy of  $g$ , but otherwise these simulations behave similarly, i.e., according to  $\langle \pi, \alpha \rangle$ .

Consider now a fixed measurement  $x$  and the measurement protocol  $\pi(x)$  over  $S \otimes R$  used to simulate it. Enumerate  $\mathcal{E}_R(X_R)$  as  $t_1, \dots, t_k$ . Since all the copies of  $x$  in  $U^{\otimes n}$  are compatible with each other, so are their images under the simulation  $d \otimes g^{\otimes n} \rightarrow f^{\otimes n}$ . When simulating different copies of  $x$  one queries a single copy of  $d$  but different copies of  $g$ , which means that whatever results one gets in  $g$  the measurements done in  $d$  are always compatible. This implies that  $\{\pi(x)(t_i) | i = 1, \dots, k\}$  is a compatible set of measurement protocols, where  $\pi(x)(t_i)$  is defined as in Remark 9. Thus they can be combined to a single measurement protocol  $P_x$  by Lemma 8. In particular, when simulating the  $i$ th copy of  $x$  we can first measure  $P_x$  obtaining an outcome  $s$ , and then perform  $\pi(x)(s)$  (which is a measurement protocol over  $R$ ) in the  $i$ th copy of  $R$ . We can then replace  $d$  by the restriction  $\hat{d}$  of  $\text{MP}(d)$  to  $\text{MP}(S)|_{\{P_x | x \in X_U\}}$ . Moreover, considering distinct

measurements  $x$  belong to different copies of  $f : U$ , we see that this set has to be a compatible set of measurement protocols. As the set  $\{P_x | x \in X_U\}$  is comeasurable, the model  $\hat{d}$  is noncontextual, so that the deterministic simulation  $\hat{d} \otimes g = \hat{d} \otimes e \otimes c \rightarrow f$  gives a probabilistic simulation  $e \rightarrow f$  as desired.

### Proof of Theorem 2

We now show how the above proof implies that there are no catalysts in the resource theory of nonlocality. Indeed, assume a simulation  $(\pi_i, \alpha_i)_{i=1}^n: d \boxtimes e \rightarrow d \boxtimes f$  in the resource theory of nonlocality. Then  $\bigotimes_{i=1}^n (\pi_i, \alpha_i)$  defines a simulation  $d \boxtimes e \rightarrow d \boxtimes f$  in the resource theory of contextuality. Pre- and postcomposing by the canonical isomorphisms

$$\left( \bigotimes_{i=1}^n S_i \right) \otimes \left( \bigotimes_{i=1}^n T_i \right) \cong \bigotimes_{i=1}^n (S_i \otimes T_i)$$

then implies that we have a simulation  $d \otimes e \rightarrow d \otimes f$ , so that the proof gives us a simulation  $(\pi', \alpha'): e \rightarrow f$ . However, we must check that this map is of the form  $\bigotimes (\pi'_i, \alpha'_i)$  in order for it to give a simulation  $e \rightarrow f$  in the resource theory of nonlocality. That this is true follows by inspecting the preceding proof. In fact, when simulating a measurement  $x$  in the  $T$  one performs a single measurement protocol in  $S$  that captures everything one might need when measuring  $x$ , and then proceeds to  $R$ . If  $x$  is a measurement at the  $i$ th measurement site, so is this “everything” and the protocol following it in  $R$ , so that the constructed simulation  $e \rightarrow f$  is indeed of the required form.

### Proof of Theorem 3

If  $\mathcal{X}$  is closed under  $\otimes$  and contains all noncontextual models, then  $\mathcal{X}$ -assisted simulation  $d \otimes e \rightarrow d \otimes f$  is given by a deterministic simulation  $d \otimes e \otimes c \rightarrow d \otimes f$  where  $c \in \mathcal{X}$ . We can then set  $g := e \otimes c$  to obtain a deterministic simulation  $d \otimes g \rightarrow d \otimes f$  and retrace the proof of Theorem 1 to obtain a deterministic simulation  $\hat{d} \otimes g = \hat{d} \otimes e \otimes c \rightarrow f$ . As  $\hat{d}$  is noncontextual, our assumptions on  $\mathcal{X}$  imply that  $\hat{d} \otimes c \in \mathcal{X}$ , so that there is an  $\mathcal{X}$ -assisted simulation  $e \rightarrow f$ .

The proof of Theorem 2 similarly gives a proof of no-catalysis for the resource theory of  $\mathcal{X}$ -assisted  $n$ -partite nonlocality.