Self-testing of physical theories, or, is quantum theory optimal with respect to some information-processing task?

Mirjam Weilenmann[1, *] and Roger Colbeck[2, †]

[1]*Institute for Quantum Optics and Quantum Information (IQOQI) Vienna,*
Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria
[2]*Department of Mathematics, University of York, Heslington, York, YO10 5DD, United Kingdom*

(Dated: August 6, 2020)

Self-testing usually refers to the task of taking a given set of observed correlations that are assumed to arise via a process that is accurately described by quantum theory, and trying to infer the quantum state and measurements. In other words it is concerned with the question of whether we can tell what quantum black-box devices are doing by looking only at their input-output behaviour and is known to be possible in several cases. Here we introduce a more general question: is it possible to self-test a theory, and, in particular, quantum theory? More precisely, we ask whether within a particular causal structure there are tasks that can only be performed in theories that have the same correlations as quantum mechanics in any scenario. We present a candidate task for such a correlation self-test and analyse it in a range of generalised probabilistic theories (GPTs), showing that none of these perform better than quantum theory. A generalisation of our results showing that all non-quantum GPTs are strictly inferior to quantum mechanics for this task would point to a new way to axiomatise quantum theory, and enable an experimental test that simultaneously rules out such GPTs.

I. INTRODUCTION

When we do experimental science we are limited to classical data in the sense that the measurement settings we choose and the outcomes we observe are always classical. However, we can use quantum systems to model how the classical outputs come about. For example, quantum theory might tell us that when a particular (classical) setting is chosen, this corresponds to the preparation of some quantum state $|\psi\rangle$. Similarly, another (classical) setting may correspond to a particular quantum measurement (say the Positive-Operator-Valued-Measure (POVM) $\{E_x\}_x$), and according to quantum theory the Born rule gives us the outcome probabilities $\langle\psi|E_x|\psi\rangle$. Often we want to go the other way. Suppose, for instance, that we wish to infer the quantum description of the state generated in a particular process. Given sufficiently many trusted measurement devices we may be able to do this using state tomography [1].

In some cases, trusted devices are not needed. For these, we can infer the quantum state and measurements (up to some symmetries) based only on the observed correlations of classical variables. This is referred to as *self-testing* [2]. For instance, by observing correlations that saturate Tsirelson's bound [3], i.e., that reach the maximal score allowed by quantum mechanics in the CHSH game [4], one can infer that the quantum state must be a singlet up to local isometries [5]. Such work was extended to GHZ states [6] in [7, 8] and subsequent work has shown that all two-qubit pure entangled states can be self-tested [9].

Self-testing as introduced above is thus the question of whether a quantum description of some devices can be inferred by only interacting with them in a classical way. Significantly, it assumes that there is a valid quantum description. In the present work we go beyond this to consider the question of what happens when this assumption is dropped. Can we, by only relying on an observed classical input-output behaviour, infer that a particular theory must be used to generate it? This is the question of self-testing of a physical theory. Of most interest to us will be self-testing of quantum theory. If we can find a task within a particular causal structure[1] that cannot be performed in any theory other than quantum mechanics, then we would have a direct test of quantum mechanics. This could also constitute an insight into the question of "why quantum theory?"

II. SELF-TESTING OF PHYSICAL THEORIES

In this paper we discuss self-testing of physical theories and a candidate task for doing so for quantum theory. Since dealing with arbitrary alternatives to quantum theory is challenging we begin the investigation by restricting our considerations to generalised probabilistic theories (GPTs) [11, 12] (i.e., *GPT self-testing*). These are a class of theories that include classical and quantum theory as well as theories with more non-local correlations still. In this framework, the state space of a system

---

* mirjam.weilenmann@oeaw.ac.at
† roger.colbeck@york.ac.uk

---

[1] In the context of the present work these are diagrams indicating the allowed flow of information during the game. For a more technical account see e.g., [10].

can be expressed as a compact convex set in a real vector space and measurements are collections of effects, linear maps from states to probabilities. Joint states can be defined for multipartite systems, and we require that these lead to non-signalling correlations.[2] In this class, there is a well-known theory that is maximally non-local with respect to bipartite correlations, referred to as *box-world*, in which PR-box correlations (i.e., those that can win the CHSH game described in Figure 1 with probability 1) are realisable.

In the present work we consider *correlation self-testing* in which we seek a causal structure and a task such that any GPT that achieves the (optimal) performance in that task must have the same set of possible correlations as quantum theory in any setting. There are several other forms of self-testing one could consider, aiming to identify, for instance, the state space of the theory, or, more generally, the states, measurements and possible transformations. We discuss these notions in our accompanying paper [13].

Given two theories $\mathcal{T}_1$ and $\mathcal{T}_2$, if there is a causal structure in which $\mathcal{T}_1$ generates some correlations that cannot be generated in $\mathcal{T}_2$, then there is a task in which $\mathcal{T}_1$ outperforms $\mathcal{T}_2$. Similarly, suppose we have a family of theories $\{\mathcal{T}_i\}_{i=1}^n$ in which for each pair there is some causal structure for which $\mathcal{T}_i$ generates some correlations that cannot be generated in $\mathcal{T}_j$. Then $n-1$ tasks are sufficient for correlation self-testing any theory within this set in the sense that by requiring a certain performance in each task, each theory can be singled out uniquely. In general we may want to consider infinite sets of theories. Correlation self-testing of quantum theory then boils down to identifying a family of tasks—optimally a single one—that are sufficient for singling it out. In this Letter, we propose a task, the adaptive CHSH game, and explain why it is a promising candidate for self-testing quantum theory.

We also note that if it is the case that quantum theory leads to a subset of the correlations of another theory, $\mathcal{T}$, in *all* causal structures, and where this inclusion is strict in some of them, then quantum theory cannot be fully correlation self-tested. Instead, our inability to find non-quantum correlations would make $\mathcal{T}$ less plausible. Furthermore, designing a self-test for $\mathcal{T}$ would lead to insights about quantum theory and point to possible additional axioms that might rule $\mathcal{T}$ out.

Before introducing the adaptive CHSH game, we make a few simple observations regarding the self-testing of other theories. Were we to observe PR-box correlations in the bipartite Bell causal structure (see Figure 1) we would be able to self-test the set of all non-signalling correlations with two possible inputs and outputs per party.
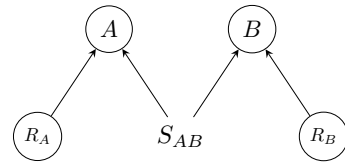

Figure 1. The bipartite Bell causal structure and CHSH game. A referee asks question R_A to Alice and R_B to Bob, choosing these uniformly. Their respective answers to these questions are A and B. These may be formed by measurement on part of a shared bipartite state S_{AB}, whose most general form depends on the theory under consideration (in quantum theory it can be any shared quantum state, for instance). In the case where A, B, R_A and R_B are binary, the CHSH game is said to be won if $A \oplus B = R_A \cdot R_B$, where $\oplus$ denotes addition modulo 2. Note that in this causal structure the measurements only act on single subsystems of the shared state.

This would correspond to a large restriction on the underlying theories to those in which PR-box correlations can be distilled.

Conversely, finding correlations achieving the quantum bound in the CHSH game eliminates theories for which these cannot occur, but does not rule out more non-local theories (e.g., those in which PR-boxes could be realised). This observation is somewhat trivial. However, it is useful for illustrating an important point: correlation self-testing of quantum theory cannot be done in any causal structure in which all the measurements can only act on single subsystems. In such causal structures, moving to a GPT that has a larger set of states cannot decrease the set of realisable correlations and hence quantum theory cannot be optimal for any task in such a causal structure (or, more precisely, if it is, it shares this optimality with all theories with larger sets of correlations).

GPTs that have larger state spaces than quantum theory have smaller effect spaces. This is because a given effect must be applicable to any of the allowed states, i.e., it must map them to a probability. Adding to the state space corresponds to additional requirements for an effect to be valid[3]. In order to find a causal structure in which quantum theory allows correlations that cannot occur within any other GPT we need to simultaneously exploit the need for a sufficiently large state space and a sufficiently large effect space.

---

[2] While non-signalling is essential so that considering causal structures makes sense, assuming local tomography is not necessary here.

[3] In a GPT, the set of allowed effects must be a subset of the dual cone to the cone of subnormalised states, so that, without further restrictions large state spaces have small effect spaces and vice versa.
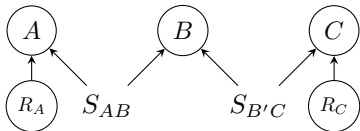
Figure 2. Causal structure of the adaptive CHSH game. The values of A, B and C as well as the referee's questions R_A and R_C determine whether the game is won. Resources are shared between A and B, and between B and C, but there are no shared tripartite resources. Although A and C can be formed by measurements on single subsystems, the value of B can be obtained by jointly measuring the B part of S_{AB} and B' part of $S_{B'C}$.

III. CORRELATION SELF-TESTING OF QUANTUM THEORY IN THE ADAPTIVE CHSH GAME

To pursue this we consider the causal structure of Figure 2 and a game played by three cooperating players, Alice, Bob and Charlie, which we call the adaptive CHSH game. A referee asks Bob to choose one of four versions of the CHSH game [4]. We call this choice B. The referee then asks Alice and Charlie questions, denoted R_A and R_C respectively, for which they have to give answers, labelled A and C respectively, where R_A, R_C, A and C can take values 0 or 1, and R_A and R_C are chosen uniformly at random by the referee. The three players win the game if Alice and Charlie's answers win the instance of the CHSH game Bob chose[4].

The instances of the CHSH game are as follows:

B (Bob's choice)	Winning condition
$(0,0)$	$(r_A \oplus 1) \cdot r_C = a \oplus c$
$(0,1)$	$(r_A \oplus 1) \cdot (r_C \oplus 1) \oplus 1 = a \oplus c$
$(1,0)$	$(r_A \oplus 1) \cdot (r_C \oplus 1) = a \oplus c$
$(1,1)$	$r_A \cdot (r_C \oplus 1) \oplus 1 = a \oplus c$

Thus, the overall winning probability of the game for a strategy that leads to a distribution $P_{ABCR_AR_C}(a,b,c,r_A,r_C)$ is

$$p_{\text{win}} = \sum_{a,b,c,r_A,r_C} P_{ABCR_AR_C}(a,b,c,r_A,r_C)Q(a,b,c,r_A,r_C),$$

where $Q(a,b,c,r_A,r_C)$ is 1 if the corresponding winning condition in the above table is met and 0 otherwise. Since R_A and R_C are chosen uniformly, we have $P_{ABCR_AR_C}(a,b,c,r_A,r_C) = \frac{1}{4}P_{ABC|r_Ar_C}(a,b,c)$.

The idea behind the use of this game is that in order to win with the highest possible probability in quantum mechanics, an entanglement swapping operation is required. If S_{AB} and $S_{B'C}$ are both maximally entangled qubit

pairs, and a Bell basis measurement is performed on the BB' systems to give the outcome B, then, by choosing the measurements generating A and C appropriately, the game can be won with probability $\frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right)$ (see our accompanying paper [13] for an explicit strategy). In order that a value as high as this can be obtained it is necessary that the same value is obtainable in the CHSH game in the bipartite Bell causal structure of Figure 1. Therefore, in any GPT for which the maximum probability of winning the CHSH game is below Tsirelson's bound of $\frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\right)$ [3], the same upper bound holds for the adaptive CHSH game. Quantum theory hence beats any theory for which Tsirelson's bound cannot be saturated.

More interestingly, in all GPTs whose joint state space is formed by taking the maximal tensor product of the single-system state space, the joint measurement on BB' can be thought of in terms of a measurement on one of the subsystems, followed by a measurement on the other (possibly depending on the result), or convex mixtures of measurements of this kind [12]. Such measurements cannot lead to any resulting non-locality between A and C, even after conditioning on the outcome, and hence in such theories the adaptive CHSH game cannot be won with probability greater than 3/4 (the maximum achievable classically). An example of such a theory is box-world, see [12, 14]. For this particular theory, previous work on *couplers* also directly implies this restriction [15–17].

From these arguments, when considering the adaptive CHSH game, quantum theory is superior to any theories in which the joint states are formed with either the minimal tensor product (where all states are separable) or the maximal tensor product. These behave like two extremes[5], with the tensor product of quantum theory sitting in between. The adaptive CHSH game is hence a reasonable candidate for a task for which quantum theory is optimal among all GPTs.

In order to show that quantum theory cannot be beaten, we need to consider GPTs whose joint state space is formed using other tensor products. However, with the exception of quantum theory[6], we are not aware of explicit alternative tensor products. Nonetheless, the single system state spaces already impose restrictions on the achievable correlations that allow us to bound their performance in the adaptive CHSH game.

To do so we consider a set of self-dual theories in which the single system state spaces take the form of a two-dimensional regular polygon with a varying number of sides, n [19, 20]. Roughly speaking, self-dual theories are those for which the set of (subnormalized) states and

---

[4] Bob's choice is not communicated to Alice and Charlie so that they are oblivious to which game they are playing (they may even give their answers before Bob).

[5] For GPTs that are locally tomographic these tensor products are known to define the minimal and the maximal joint state spaces that are possible.

[6] The same is true for quantum theory over the real numbers, which leads to the same correlations as quantum theory everywhere [18].

effects are the same (for instance, in quantum theory, subnormalized states are positive operators with trace at most 1, as are the POVM elements, which correspond to effects). In the case of odd n, taking the effect space to be the maximal possible, i.e., the set of all linear maps from states to probabilities, naturally yields a self-dual theory, while for even n one has to transform these spaces (see [20] and our accompanying paper [13]). Our main result for these GPTs is the following.

Theorem (informal): Consider any GPT where the single-system state space is a self-dualized regular polygon system with up to 30 extremal states and the joint state space is any tensor product between the minimal and the maximal. If the sources S_{AB} and $S_{B'C}$ each correspond to one pair of such gbits then the winning probability in the adaptive CHSH game is upper bounded by the quantum value.

To prove this we use the fact that the winning probability in the adaptive CHSH game can be upper bounded by the maximum probability of winning the regular CHSH game. Using linear programming we evaluated the latter using any bipartite state from the maximal tensor product of the considered state spaces (for other tensor products these are upper bounds). The values obtained are illustrated in Figure 3, and we found analytic formulae that appear to repeat modulo 8, each of which tends to the quantum value as n tends to infinity. For n mod $8 \neq 0$ the values are strictly smaller than the maximal quantum value, while for n mod $8 = 0$ the upper bound is equal to the quantum value. However, we do not expect any of these upper bounds to be achievable in the adaptive CHSH game because they are based on the maximal tensor product in which case we know that the adaptive CHSH game cannot be won with probability larger than 3/4.

Thus, the above theorem is expected to generalize to any n and to a statement that the winning probabilities are strictly smaller than the quantum one even for n mod $8 = 0$. Further details of these results as well as a discussion of bipartite polygon state spaces without self-duality can be found in our accompanying paper [13].

A number of challenges need to be overcome to extend our results to a full correlation self-testing of quantum theory. Firstly, there is a wide variety of possible theories, and not many theoretical results have been developed that can simultaneously apply to them all. Secondly, we lack general bounds on when non-locality can be distilled, and by how much. For instance it could be that individual systems in a particular GPT always give rise to correlations between A and C that cannot violate Tsirelson's bound. However, if, given many copies of such a system, we could distil correlations that violate Tsirelson's bound without communication, then the theory could win the adaptive CHSH game better than quantum theory. Thirdly, there is a lack of known ways to construct joint state spaces other than with the minimal and maximal tensor products, both of which we know do not give an advantage in this game. Thus, the question
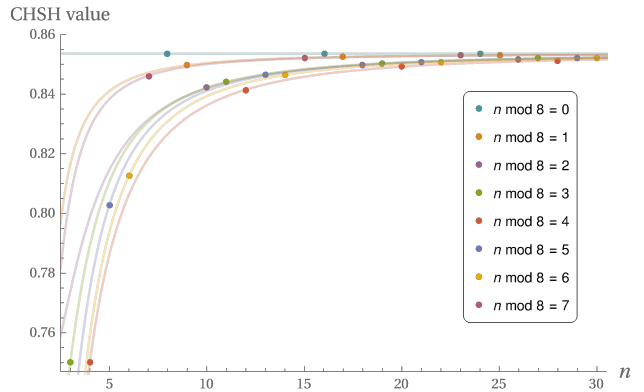

Figure 3. Maximal winning probability in the CHSH game achievable in the maximal tensor product of self-dualized polygon systems. Along the horizontal axis we display the number of extremal vertices n of the local state space, while on the vertical axis we display the respective winning probability. The points are the values obtained in our optimizations while the curves depict respective formulae these values follow. The colours are used to group points that appear to follow the same analytic formula.

we have asked necessitates the development of several lines of research before it can be answered in full generality. In an accompanying paper [13] we report additional progress in this direction.

IV. RELATED WORK

The question "why quantum mechanics?" has been a topic of debate since the conception of the theory. Unlike other theories that have firm physical principles behind them (such as the principle that the laws of physics are the same in every reference frame for special relativity), quantum mechanics is usually presented as a series of mathematical axioms whose underlying physical significance is unclear. There have been numerous attempts to give quantum mechanics a more physical axiomatisation, going back to Popescu and Rohrlich [21] who asked whether quantum mechanics is the most non-local theory that obeys the no-signalling principle. That it is not follows because PR-boxes are non-signalling but can win the CHSH game of Figure 1 with probability 1, in violation of Tsirelson's bound [3] (the PR-box correlations and the impossibility of realising them in quantum theory was also realised by Tsirelson [22]).

There have been several other attempts to find such a principle from which quantum theory naturally follows [23–27]. Each of these principles imposes a restriction on the correlations a 'reasonable' theory may produce, however, none of them singles out the set of quantum correlations exactly. Instead, they are also obeyed by correlations that are not achievable quantum mechanically, in particular by the set of *almost quantum cor-*

relations (potentially with the exception of Information Causality, for which numerical evidence supports this, but it is not strictly proven) [28]. We also remark that establishing optimality with respect to a task, as considered in this paper, is more objective than deciding whether axioms seem 'reasonable'.

V. CONCLUSIONS

We have discussed the possibility of showing optimality of quantum theory using the adaptive CHSH game, but there are many other tasks that could be considered instead, and it could be the case that a family of tasks are required. In particular, our game only involves sharing bipartite systems, whereas consideration of more parties may be necessary to single out quantum mechanics. In addition, although we have phrased the task via a game, we could alternatively look to find a set of correlations achievable in a causal structure within quantum theory but not in any other GPT. The triangle causal structure [29] is a candidate for this because it also has the possibility of joint measurements on multiple subsystems.

If it is possible to show that the optimal performance in the adaptive CHSH game (or an alternative) is only achieved by theories with the same correlations as quantum theory, this will point towards a new way to axiomatise quantum theory. Admittedly, an axiom of the form "the theory is the one with the highest winning probability with respect to the adaptive CHSH game" would not be especially natural. However, there may be a deeper and more natural principle underlying it. More significantly, attempting to experimentally observe the correlations that optimally win the game would be a way to directly rule out a range of alternative theories, hence providing a strong confirmation of the validity of quantum mechanics. This could have impact on the search for future theories, such as candidates for uniting quantum mechanics and gravity, although at the moment this remains speculative. Nevertheless, we consider this a promising research direction that has the potential to be transformative.

ACKNOWLEDGMENTS

[1] U. Fano, "Description of states in quantum mechanics by density matrix and operator techniques," Reviews of Modern Physics **29**, 74–93 (1957).

[2] D. Mayers and A. Yao, "Quantum cryptography with imperfect apparatus," in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS-98)* (IEEE Computer Society, Los Alamitos, CA, USA, 1998) pp. 503–509.

[3] B. S. Cirel'son, "Quantum generalizations of Bell's inequality," Letters in Mathematical Physics **4**, 93–100 (1980).

[4] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," Physical Review Letters **23**, 880–884 (1969).

[5] S. Popescu and D. Rohrlich, "Which states violate Bell's inequality maximally?" Physics Letters A **169**, 411–414 (1992).

[6] D. M. Greenberger, M. Horne, and A. Zeilinger, "Going beyond Bell's theorem," in *Bell's Theorem, Quantum Mechanics and Conceptions of the Universe*, edited by M. Kafatos (Kluwer Academic, Dordrecht, The Netherlands, 1989) pp. 69–72.

[7] R. Colbeck, *Quantum and Relativistic Protocols For Secure Multi-Party Computation*, Ph.D. thesis, University of Cambridge (2007), also available as arXiv:0911.3814.

[8] R. Colbeck and A. Kent, "Private randomness expansion with untrusted devices," Journal of Physics A **44**, 095305 (2011).

[9] A. Coladangelo, K. T. Goh, and V. Scarani, "All pure bipartite entangled states can be self-tested," Nature Communications **8**, 15485 (2017).

[10] M. Weilenmann and R. Colbeck, "Analysing causal structures with entropy," Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **473**, 20170483 (2017).

[11] L. Hardy, "Quantum theory from five reasonable axioms," (2001), arXiv:quant-ph/0101012 [quant-ph].

[12] J. Barrett, "Information processing in generalized probabilistic theories," Physical Review A **75**, 032304 (2007).

[13] M. Weilenmann and R. Colbeck, "Toward correlation self-testing of quantum theory in the adaptive Clauser-Horne-Shimony-Holt game," Physical Review A **102**, 022203 (2020).

[14] D. Gross, M. Müller, R. Colbeck, and O. C. O. Dahlsten, "All reversible dynamics in maximally nonlocal theories are trivial," Physical Review Letters **104**, 080402 (2010).

[15] A. J. Short, S. Popescu, and N. Gisin, "Entanglement swapping for generalized nonlocal correlations," Physical Review A **73**, 012101 (2006).

[16] P. Skrzypczyk, N. Brunner, and S. Popescu, "Emergence of quantum correlations from nonlocality swapping," Physical Review Letters **102**, 110402 (2009).

[17] P. Skrzypczyk and N. Brunner, "Couplers for non-locality swapping," New Journal of Physics **11**, 073014 (2009).

[18] M. McKague, M. Mosca, and N. Gisin, "Simulating quantum systems using real Hilbert spaces," Physical Review Letters **102**, 020505 (2009).

[19] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, "Limits on nonlocal correlations from the structure of the local state space," New Journal of Physics **13**, 063024 (2011).

[20] P. Janotta and R. Lal, "Generalized probabilistic theories without the no-restriction hypothesis," Physical Review A **87**, 052131 (2013).

[21] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," Foundations of Physics **24**, 379–385 (1994).

[22] B. S. Tsirelson, "Some results and problems on quantum Bell-type inequalities," Hadronic Journal Supplement **8**, 329–345 (1993).

[23] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, "Limit on nonlocality in any world in which communication complexity is not trivial," Physical Review Letters **96**, 250401 (2006).

[24] N. Linden, S. Popescu, A. J. Short, and A. Winter, "Quantum nonlocality and beyond: Limits from nonlocal computation," Physical Review Letters **99**, 180502 (2007).

[25] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, "Information causality as a physical principle," Nature **461**, 1101–1104 (2009).

[26] M. Navascues and H. Wunderlich, "A glance beyond the quantum model," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **466**, 881–890 (2010).

[27] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acín, "Local orthogonality as a multipartite principle for quantum correlations," Nature Communications **4**, 2263 (2013).

[28] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, "Almost quantum correlations," Nature communications **6**, 6288 (2015).

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