What can Hardy-type proofs of nonlocality teach us about causal inference?

RWS*
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5
(Dated: 22/07/14)

* rspekkens@perimeterinstitute.ca

In Ref. [1], the standard proof of Bell's theorem is presented in the language of causal inference. In particular, the CHSH inequality emerges as a special case of what Pearl calls an "instrumental inequality". Hardy's proof of Bell's theorem is quite different from the standard proof and the following question naturally arises: is there a generic tool for classical causal inference of which the Hardy argument can be considered a special case when applied to the M-shaped causal structure of the Bell experiment?

To try and answer this question, we apply Hardy-type reasoning to the triangle causal structure, that is, the one with three observed variables, each pair of which have a common cause. We show that this sort of reasoning does indeed facilitate causal inference in the case of the triangle causal structure, thereby lending some evidence to the notion that this style of argument has the potential to be generalized into a generic tool for classical causal inference.

I. RECASTING THE HARDY PARADOX IN THE LANGUAGE OF CAUSAL MODELS

We begin by recalling the Hardy paradox. For this purpose, we reproduce a discussion found in Ref. [2].

> In outcome-deterministic ontological models that are local or noncontextual, implications among value assignments of observables are transitive because these value assignments do not depend on the context (local or remote) of the measurement. The failure of the transitivity of implication therefore implies the impossibility of such models. Again, we find that this conclusion has been reached before in the literature on nonlocality. Specifically, the Hardy-type proof of nonlocality [3] can be expressed in this fashion [4], a fact that was first noted by Stapp [5] (for a simplified account, see Refs. [6, 7]).
>
> We begin by presenting Hardy's proof of nonlocality in its standard form. It uses a pair of binary-outcome observables on each wing of the experiment. Hardy demonstrated a way of choosing these observables such that for any partially entangled pure state, the correlations between these observables satisfy:
>
> $$A_1 = 1 \implies B_1 = 1, \tag{1}$$
> $$B_2 = 1 \implies A_2 = 1, \tag{2}$$
>
> while
>
> $$\text{sometimes } (A_1 = 1 \text{ and } B_2 = 1) \tag{3}$$
>
> (i.e. with probability $p_{\text{Hardy}} \equiv p(A_1 = 1 \text{ and } B_2 = 1) > 0$), and
>
> $$\text{never } (A_2 = 1 \text{ and } B_1 = 1)\,. \tag{4}$$
>
> We can express this as a failure of the transitivity of implication as follows. From Eqs. (1), (4) and (2) (in its contrapositive form), we infer respectively,
>
> $$A_1 = 1 \implies B_1 = 1, \tag{5}$$
> $$B_1 = 1 \implies A_2 = 0, \tag{6}$$
> $$A_2 = 0 \implies B_2 = 0. \tag{7}$$
>
> which we summarize graphically by
>
> $$\begin{array}{ccc} A_1 = 1 & \Longrightarrow & B_1 = 1 \\ & \swarrow\!\!\!\swarrow & \\ A_2 = 0 & \Longrightarrow & B_2 = 0 \end{array}$$
>
> If transitivity held, then these three inferences would imply that
>
> $$A_1 = 1 \implies B_2 = 0. \tag{8}$$
>
> However, this contradicts Eq. (3) and consequently transitivity must fail. More explicitly, taking $\implies$ to be material implication, the negation of Eq. (8) is the conjunction of $A_1 = 1$ and $B_2 = 1$,
>
> $$\neg\,(A_1 = 1 \implies B_2 = 0) = (A_1 = 1 \text{ and } B_2 = 1)\,, \tag{9}$$

so that the probability $p_{\text{Hardy}} \equiv p(A_1 = 1 \text{ and } B_2 = 1)$ quantifies the frequency with which the transitivity of implication fails.

We now consider the status of this sort of proof for the PR box. By relabeling the outcomes of the standard PR box, one can obtain correlations of the form

$$A_1 = B_1 \tag{10}$$
$$A_1 = B_2 \tag{11}$$
$$A_2 = B_1 \oplus 1 \tag{12}$$
$$A_2 = B_2, \tag{13}$$

with marginals of the form $p(A_1 = 0) = p(A_2 = 0) = p(B_1 = 0) = p(B_2 = 0) = 1/2$. Eqs. (10), (12) and (13) imply the inferences of Eqs. (5), (6), and (7) respectively. Meanwhile, Eq. (11), together with the fact that $p(A_1 = 1) = 1/2$, implies that sometimes $A_1 = 1$ and $B_2 = 1$, or equivalently, that sometimes Eq. (8) fails, so that we have a contradiction with transitivity. Indeed, the probability of this occurring is $p_{\text{Hardy}} = p(A_1 = 1 \text{ and } B_2 = 1) = 1/2$.

Actually, p_{Hardy} only quantifies the probability for one particular kind of contradiction, which requires $A_1 = 1$ to get going. In the rest of the cases, where $A_1 = 0$, we still obtain a contradiction because Eqs. (10), (12) and (13) also imply inferences of the form of Eqs. (5), (6), and (7) where $A_a \Leftrightarrow A_a \oplus 1$ and $B_b \Leftrightarrow B_b \oplus 1$. Transitivity then implies that $A_1 = 0 \implies B_2 = 1$, while Eq. (11) contradicts this. So one obtains a contradiction with certainty for the PR box.

We now reformulate the Hardy argument, where instead of using the assumption of local causality (or the assumption of the transitivity of implication that it implies), we use the assumption that the causal structure is the M-shaped structure associated to the Bell experiment, depicted in Fig. 1

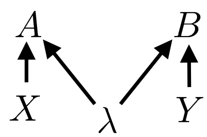

FIG. 1. The causal structure on which Bell's theorem is based.

The outcome and setting on the left are labelled A and X, the outcome and the setting on the right are labelled B and Y, while the common cause hidden variable is labelled λ. Note, that we will deviate from the notation used in the quote given above by denoting the two settings by $X \in \{0, 1\}$ rather than $X \in \{1, 2\}$.

One can cast the Hardy proof as follows:

Proposition 1 *For the causal structure of Fig. 1, the following inference holds*

$$\text{if } \; p(A = 1, B = 1|X = 0, Y = 0) > 0 \tag{14}$$
$$\text{and } \; p(B = 1|Y = 1, A = 1, X = 0) = 1 \tag{15}$$
$$\text{and } \; p(A = 1|X = 1, B = 1, Y = 0) = 1 \tag{16}$$
$$\text{then } \; p(A = 1, B = 1|X = 1, Y = 1) > 0. \tag{17}$$

As Hardy has shown, there exist quantum correlations that violate this inference (i.e. satisfy the antecedents, (14)-(16) but violate the consequent (17)). It follows that these quantum correlations cannot be explained by the causal structure of Fig. 1.

Proof. From the causal structure, we can infer that the LHS of Eq. (15) can be expressed as

$$p(B = 1|Y = 1, A = 1, X = 0) = \sum_{\lambda} p(B = 1|Y = 1, \lambda)p(\lambda|A = 1, X = 0). \tag{18}$$

From Eq. (15) it follows that

$$\forall \lambda : p(\lambda|A = 1, X = 0) > 0, \text{ we have } p(B = 1|Y = 1, \lambda) = 1. \tag{19}$$

Similarly, from the causal structure, we can infer that the LHS of Eq. (16) can be expressed as

$$p(A = 1|X = 1, B = 1, Y = 0) = \sum_{\lambda} p(A = 1|X = 1, \lambda)p(\lambda|B = 1, Y = 0). \tag{20}$$

and from Eq. (16) it then follows that

$$\forall\lambda : p(\lambda|B=1, Y=0) > 0, \text{ we have } p(A=1|X=1,\lambda) = 1. \tag{21}$$

Finally, from the causal structure, we can infer that the LHS of Eq. (14) can be expressed as

$$p(A=1, B=1|X=0, Y=0) = \sum_{\lambda} p(A=1|X=0,\lambda)p(\lambda|B=1|Y=0,\lambda)p(\lambda). \tag{22}$$

It then follows from Eq. (14) that there must exist a value of λ, which we denote λ_* such that

$$p(\lambda_*) > 0, \tag{23}$$
$$p(A=1|X=0,\lambda_*) > 0, \tag{24}$$
$$p(B=1|Y=0,\lambda_*) > 0. \tag{25}$$

By Bayesian inversion, we obtain

$$p(\lambda_*|A=1, X=0) = \frac{p(A=1|X=0,\lambda_*)p(\lambda_*)}{p(A=1)} > 0, \tag{26}$$
$$p(\lambda_*|B=1, Y=0) = \frac{p(B=1|Y=0,\lambda_*)p(\lambda_*)}{p(B=1)} > 0. \tag{27}$$

It then follows from Eqs (19) and (21) that

$$p(B=1|Y=1,\lambda_*) = 1 \tag{28}$$
$$p(A=1|X=1,\lambda_*) = 1. \tag{29}$$

Finally, from the causal structure, we can infer that the LHS of Eq. (17) can be expressed as

$$p(A=1, B=1|X=1, Y=1) = \sum_{\lambda} p(A=1|X=1,\lambda)p(B=1|Y=1,\lambda)p(\lambda). \tag{30}$$

Making use of Eqs. (23), (28) and (29), we obtain Eq. (17). ∎

II. THE TRIANGLE CAUSAL STRUCTURE

We now consider the causal structure depicted in Fig. 2.

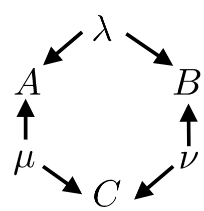

FIG. 2. The triangle causal structure.

The three observed variables are denoted A, B and C, while λ denotes the common cause of A and B, μ denotes the common cause of A and C, and ν denotes the common cause of B and C.

Proposition 2 *The causal structure of Fig. 2 implies the following inference*

$$\text{if } \; p(A=1) > 0 \tag{31}$$
$$\text{and } \; p(B=1) > 0 \tag{32}$$
$$\text{and } \; p(C=1) > 0 \tag{33}$$
$$\text{and } \; A=1 \implies B=0 \tag{34}$$
$$\text{and } \; B=1 \implies C=0 \tag{35}$$
$$\text{and } \; C=1 \implies A=0 \tag{36}$$
$$\text{then } \; p(A=0, B=0, C=0) > 0. \tag{37}$$

We can express this in terms of conditional probabilities as follows

$$\text{if}\;\; p(A=1)>0 \tag{38}$$
$$\text{and}\;\; p(B=1)>0 \tag{39}$$
$$\text{and}\;\; p(C=1)>0 \tag{40}$$
$$\text{and}\;\; p(B=0|A=1)=1 \tag{41}$$
$$\text{and}\;\; p(C=0|B=1)=1 \tag{42}$$
$$\text{and}\;\; p(A=0|C=1)=1 \tag{43}$$
$$\text{then}\;\; p(A=0,B=0,C=0)>0. \tag{44}$$

Given the inference, one can show that the distribution

$$p(A,B,C)=\frac{1}{3}\left([001]+[010]+[100]\right). \tag{45}$$

is inconsistent with the triangle causal structure. The reason is that this distribution satisfies the antecedent of the inference, Eqs. (38)-(40) and Eqs. (41)-(43), and yet it violates the consequent of the inference, Eq. (44).

It is an open question whether this distribution can be realized quantum mechanically or in a generalized probabilistic theory.

Proof. Given the causal structure, the LHS of Eq. (41) can be expressed as

$$p(B=0|A=1)=\sum_{\lambda,\nu}p(B=0|\lambda,\nu)p(\lambda|A=1)p(\nu). \tag{46}$$

Then from Eq. (41), it follows that

$$\forall\nu\forall\lambda: p(\nu)>0,\; p(\lambda|A=1)>0,\;\; \text{we have}\; p(B=0|\lambda,\nu)=1. \tag{47}$$

Using this same pattern of argument for each of the three inferences (41)-(43), we get the following three results:

$$\forall\nu\forall\lambda: p(\nu)>0,\; p(\lambda|A=1)>0,\;\; \text{we have}\; p(B=0|\lambda,\nu)=1, \tag{48}$$
$$\forall\mu\forall\nu: p(\mu)>0,\; p(\nu|B=1)>0,\;\; \text{we have}\; p(C=0|\nu,\mu)=1, \tag{49}$$
$$\forall\lambda\forall\mu: p(\lambda)>0,\; p(\mu|C=1)>0,\;\; \text{we have}\; p(A=0|\lambda,\mu)=1, \tag{50}$$
$$\tag{51}$$

Given the causal structure, we can express $p(A=1)$ as

$$p(A=1)=\sum_{\lambda,\mu}p(A=1|\lambda,\mu)\;p(\lambda,\mu). \tag{52}$$

Combining this with the fact that $p(A=1)>0$, Eq. (38), it follows that there exist values of λ and μ, which we denote λ_* and μ_0, such that

$$p(\lambda_*,\mu_0)>0 \tag{53}$$
$$p(A=1|\lambda_*,\mu_0)>0. \tag{54}$$

From the causal structure, we infer that

$$p(\lambda_*,\mu_0)=p(\lambda_*)p(\mu_0), \tag{55}$$

so that we have

$$p(\lambda_*)>0 \tag{56}$$
$$p(\mu_0)>0. \tag{57}$$

It also follows from the causal structure that

$$p(A=1|\lambda_*)=\sum_{\mu}p(A=1|\lambda_*,\mu)p(\mu), \tag{58}$$

and consequently

$$p(A=1|\lambda_*) > 0. \tag{59}$$

By Bayesian inversion, we have

$$p(\lambda_*|A=1) = \frac{p(A=1|\lambda_*)p(\lambda_*)}{p(A=1)} > 0. \tag{60}$$

Finally, recalling Eq. (48), we can infer that there exists a λ_* such that

$$\begin{aligned} &p(\lambda_*) > 0, \\ &\forall \nu : p(\nu) > 0 \;\text{ we have } p(B=0|\lambda_*, \nu) = 1. \end{aligned} \tag{61}$$

By similar arguments, we can derive that there exists a μ_* such that

$$\begin{aligned} &p(\mu_*) > 0, \\ &\forall \lambda : p(\lambda) > 0 \;\text{ we have } p(C=0|\lambda, \mu_*) = 1. \end{aligned} \tag{62}$$

and there exists a ν_* such that

$$\begin{aligned} &p(\nu_*) > 0, \\ &\forall \mu : p(\mu) > 0 \;\text{ we have } p(A=0|\mu, \nu_*) = 1. \end{aligned} \tag{63}$$

Finally, given the causal structure, we can express the LHS of Eq. (44) as

$$p(A=0, B=0, C=0) = \sum_{\lambda,\mu,\nu} p(A=0|\lambda,\mu)p(B=0|\lambda,\nu)p(C=0|\mu,\nu)p(\lambda)p(\mu)p(\nu). \tag{64}$$

In particular, this implies that

$$p(A=0, B=0, C=0) \geq p(A=0|\lambda_*,\mu_*)p(B=0|\lambda_*,\nu_*)p(C=0|\mu_*,\nu_*)p(\lambda_*)p(\mu_*)p(\nu_*).$$

It then follows from Eqs. (61), (62) and (63) that

$$\begin{aligned} p(A=0, B=0, C=0) &\geq p(\lambda_*)p(\mu_*)p(\nu_*) \\ &> 0 \end{aligned} \tag{65}$$

Therefore, we have proven Eq. (44). ■

III. HARDY-TYPE INFERENCES CAN BE STRONGER THAN ALL ENTROPIC INEQUALITIES

In Causal Inference there is a well-established technique for deriving entropic inequalities implied by a given causal structure [8, 9]. For the causal structure associated with Bell's theorem in Fig. 1, for example, there is a well known entropic Bell inequality [9, 10]. Like its corresponding probabilistic Bell inequality, the entropic variant is is capable of excluding many distributions which one could not exclude by considering only conditional independence relations. Indeed, the enumeration of all entropic inequalities which follow from a given causal structure is currently the state-of-the-art method for capturing implications from the causal structure beyond conditional independence relations [11].

However, the method of entropic inequalities is not completely sufficient. For at least some causal structures, there are disallowed probability distributions which manage to evade detection when tested against the complete set of entropic inequalities. Of course, if a distribution passes all the entropic inequalities, it usually takes some ingenuity to prove that it is nevertheless incompatible with the causal structure. The triangle scenario of Fig. 2 is a causal structure who's complete set of entropic inequalities is known to be insufficient [12]. However, the question of the sufficiency of the complete set of entropic inequalities was not answered for the case where the three observed variables in the triangle are binary. We answer it here in the negative by showing that the distibution in Eq. (45) satisfies all entropic inequalities, but is nevertheless disallowed perusant to the Hardy-type inference of **Prop. 2**.

In the triangle scenario there are precisely three unique fundamental entropic inequalities, under the obvious symmetry of relabelling the three observed variables. They are

$$H(A) \geq I(A:B) + I(A:C) \tag{66}$$
$$H(A,B) \geq I(A:B:C) + I(A:B) + I(A:C) + I(B:C) \tag{67}$$
$$\frac{H(A)+H(B)+H(C)}{2} \geq I(A:B:C) + I(A:B) + I(A:C) + I(B:C) \tag{68}$$

where bipartite mutual information may be understood as $I(A:B) \equiv H(A)+H(B)-H(A,B)$ and tripartite mutual information is defined as $I(A:B:C) \equiv H(A)+H(B)+H(C)-H(A,B)-H(A,C)-H(B,C)+H(A,B,C)$. An entropy is a weighted logarithmic sum of probabilities, for binary variables

$$\begin{aligned}
H(A,B,C) &= \sum_{a,b,c=0}^{1} \begin{cases} p(A=a,B=a,C=c)\log\left(\frac{1}{p(A=a,B=a,C=c)}\right) & p(A=a,B=a,C=c)>0 \\ 0 & p(A=a,B=a,C=c)=0 \end{cases} \\
H(A,B) &= \sum_{a,b=0}^{1} \begin{cases} p(A=a,B=a)\log\left(\frac{1}{p(A=a,B=a)}\right) & p(A=a,B=a)>0 \\ 0 & p(A=a,B=a)=0 \end{cases} \\
H(A) &= \sum_{a=0}^{1} \begin{cases} p(A=a)\log\left(\frac{1}{p(A=a)}\right) & p(A=a)>0 \\ 0 & p(A=a)=0 \end{cases}.
\end{aligned} \tag{69}$$

The base of the logarithm is conventionally taken to be 2, but the base is irrelevant for the purpose of evaluating entropic inequalities. Applied to the symmetric distribution of Eq. (45) we find

$$H(A,B,C) = H(A,B) = \log(3) \text{ and } H(A) = \log\left(3\times 2^{-2/3}\right) \tag{70}$$

From these definitions it is straightforward to verify that the W-type distribution of Eq. (45) respects Eqs. (66)-(68), thus illustrating the advantage of Hardy-type inference over entropic inequalities in terms of recognizing this distribution as incompatible with the triangle scenario.

IV. FROM INFERENCES CONCERNING OBSERVED VARIABLES TO INEQUALITIES

I suspect that using the same techniques that Ravi Kunjwal and I have used for deriving noncontextuality inequalities, we can replace the inference in proposition 1 with an inequality. Perhaps something like

$$\begin{aligned}
&p(A=1,B=1|X=1,Y=1) \\
&\geq p(B=1|Y=1,A=1,X=0)\,p(A=1|X=1,B=1,Y=0)\,p(A=1,B=1|X=0,Y=0).
\end{aligned} \tag{71}$$

I will need to think about this more.

A different approach would be to look to the pre-existing literature that seeks to derive Bell inequalities from Hardy-type reasoning. It appears that many groups have considered this question [13–17] (it is unclear whether this list of references is complete).

The idea is that perhaps one of these approaches can be generalized such that we can derive inequalities for the triangle causal structure.

I will may a small start on this problem here by considering the approach of Ghirardi and Marinotta. They state that: "we have shown that the existing Hardy-like nonlocality without inequalities proofsfor bipartite states [...] are simply particular instances of violations of the Clauser-Horne inequality." If we believe their claim, then it implies that turning the Hardy-type reasoning about transitivity of inferences into an inequality results in precisely the CH inequality, which we can summarize as follows.

Proposition 3 *For the Bell-scenario causal structure of Fig. 1, the following inequality holds*

$$\begin{aligned}
0 \leq\, &p(A=1,B=1|X=0,Y=0) \\
&-p(A=1,B=1|X=0,Y=1) \\
&-p(A=1,B=1|X=1,Y=0) \\
&-p(A=1,B=1|X=1,Y=1) \\
&+p(A=1|X=1)+p(B=1|Y=1) \leq 1.
\end{aligned} \tag{72}$$

Their argument relies on looking at the measures of various sets. Inspired by the details of their proof, can we use similar reasoning in the context of the triangle causal structure to get some inequalities?

The Ghirardi and Marinotta argument looks similar to the way that Tobias describes how he derived some inequalities for the triangle causal structure. I suspect, therefore, that this approach might just lead to the inequalities that Tobias has already found. Even if it did, it would be interesting to draw the parallel with the Hardy paradox situation.

It might be worth looking at some of the other papers in the list I provided to see if the reasoning there can be ported over to the triangle causal structure.

Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation.

[1] C. J. Wood and R. W. Spekkens, arXiv:1208.4119 (2012).
[2] Y.-C. Liang, R. W. Spekkens, and H. M. Wiseman, Phys. Rep. **506**, 1 (2011).
[3] L. Hardy, Phys. Rev. Lett. **71**, 1665 (1993).
[4] D. Boschi, S. Branca, F. De Martini, and L. Hardy, Phys. Rev. Lett. **79**, 2755 (1997).
[5] H. P. Stapp, in *Mind, Matter and Quantum Mechanics*, The Frontiers Collection (Springer Berlin Heidelberg, 2009) pp. 281–300.
[6] N. D. Mermin, Am. J. Phys. **66**, 920 (1998).
[7] W. G. Unruh, quant-ph/9710032 (1997).
[8] T. Fritz and R. Chaves, IEEE Trans. Info. Theo. **59**, 803 (2013).
[9] R. Chaves, L. Luft, and D. Gross, New J. Phys. **16**, 043001 (2014).
[10] R. Chaves and T. Fritz, Phys. Rev. A **85**, 032113 (2012).
[11] J. Henson, R. Lal, and M. F. Pusey, New J. Phys. **16**, 113043 (2014).
[12] T. Fritz, New J. Phys. **14**, 103001.
[13] N. D. Mermin, Am. J. Phys. **62**, 880 (1994).
[14] A. Garuccio, Phys. Rev. A **52**, 2535 (1995).
[15] G. Ghirardi and L. Marinatto, Phys. Lett. A **372**, 1982 (2008).
[16] D. Braun and M.-S. Choi, Phys. Rev. A **78**, 032114 (2008).
[17] L. Mancinska and S. Wehner, arXiv:1407.2320 (2014).
[18] R. Chaves, C. Majenz, and D. Gross, Nat. Comm. **6**, 5766 (2015).