A quantum advantage for inferring causal structure

Katja Ried[1,2,3]*[†], Megan Agnew[1,2][†], Lydia Vermeyden[1,2], Dominik Janzing[4], Robert W. Spekkens[3] and Kevin J. Resch[1,2]

The problem of inferring causal relations from observed correlations is relevant to a wide variety of scientific disciplines. Yet given the correlations between just two classical variables, it is impossible to determine whether they arose from a causal influence of one on the other or a common cause influencing both. Only a randomized trial can settle the issue. Here we consider the problem of causal inference for quantum variables. We show that the analogue of a randomized trial, causal tomography, yields a complete solution. We also show that, in contrast to the classical case, one can sometimes infer the causal structure from observations alone. We implement a quantum-optical experiment wherein we control the causal relation between two optical modes, and two measurement schemes—with and without randomization—that extract this relation from the observed correlations. Our results show that entanglement and quantum coherence provide an advantage for causal inference.

The slogan 'correlation does not imply causation' is meant to capture the fact that any joint probability distribution over two variables can be explained not only by a causal influence of one variable on the other, but also by a common cause acting on both[1]. We here investigate whether a similar ambiguity holds for quantum systems, and we show that, surprisingly, it does not.

Finding causal explanations of observed correlations is a fundamental problem in science, with applications ranging from medicine and genetics to economics[2,3]. As a practical illustration, consider a drug trial. Naively, a correlation between the variables treatment and recovery may suggest a causal influence of the former on the latter. But suppose men are more likely than women to seek treatment, and also more likely to recover spontaneously, regardless of treatment. In this case, gender is a common cause, inducing correlations between treatment and recovery even if there is no cause–effect relation between them.

Unless one can make strong assumptions about the nature of the causal mechanisms[4], the only way to distinguish between the two possibilities is to replace observation of the early variable with an intervention on it. For instance, pharmaceutical companies do not leave the choice of treatment to the subjects of their trials, but carefully randomize the assignment of drug or placebo. This ensures that the administered treatment is statistically independent of any potential common causes with recovery. Consequently, any correlations with recovery that persist herald a directed causal influence. The question of whether there were in fact common causes can be answered by tracking whether recovery correlates with the subjects' intent to treat. Thus, the ability to intervene allows a complete solution of the causal inference problem: it reveals both which variables are causes of which others and, via the strength of the correlations, the precise mathematical form of the causal dependencies.

In this article, we consider the quantum version of this causal inference problem. The challenge is to infer, by probing the correlations between two temporally ordered quantum systems, whether one causally influences the other, whether a common cause acts on both, or whether a combination of the two possibilities

holds. Despite the restrictions that quantum theory places on learning about systems—that not all observables that can be defined of a system can be measured precisely at the same time—we show that the ability to intervene on the early quantum system allows a complete solution of the problem. This constitutes a new type of tomography, which subsumes tomography of bipartite states and tomography of processes, and promises applications for determining whether the state evolution implemented by a given device is Markovian. We implement this new type of tomography experimentally and obtain a complete description of the causal structure.

The real surprise, however, is that even if one only has the ability to observe the early system, the quantum correlations hold signatures of the causal structure—in other words, certain types of correlation do imply causation. In a recent paper, Fitzsimons, Jones and Vedral[5] defined a function of the observed correlations which acts as a witness of causal influence, by ruling out a purely common-cause explanation. We here present a larger framework that places this result on an equal footing with an analogous result for witnessing common-cause relations. We exploit these facts to devise, for a particular class of causal scenarios, a complete solution of the causal inference problem using observation alone—a task that is impossible classically. We implement a family of such causal scenarios experimentally and confirm our conclusion.
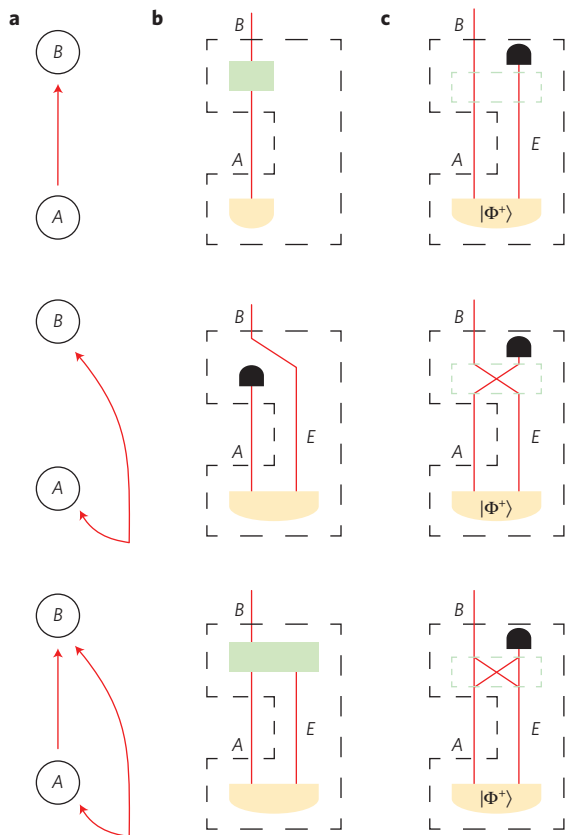
The quantum causal inference problem

The two quantum systems whose causal structure we are probing will be denoted A and B, with A preceding B in time. The dynamics relating them may be arbitrarily complicated, involving any number of additional systems and interactions. Nonetheless, any non-trivial causal relation that is induced between A and B takes one of three forms: A could be a cause of B, the two could be influenced by a common cause, or there could be a mixture of the two causal mechanisms (either a probabilistic mixture or a case where both act simultaneously). The three possibilities are depicted in Fig. 1a,b.

---

[1]Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. [2]Department of Physics & Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada. [3]Perimeter Institute for Theoretical Physics, 31 Caroline St. N, Waterloo, Ontario N2L 2Y5, Canada. [4]Max Planck Institute for Intelligent Systems, Spemannstraße 38, 72076 Tübingen, Germany. [†]These authors contributed equally to this work. *e-mail: kried@perimeterinstitute.ca

Figure 1 | The quantum causal inference problem. We aim to discriminate the three possible causal relations that may hold between a pair of temporally ordered quantum systems: (top to bottom) cause–effect, common cause or a combination of both. **a**, Directed acyclic graphs, where nodes represent quantum systems and directed edges represent causal influences, are the conventional depiction of causal structure in the causal inference literature[2,3]. **b**, Quantum circuits depicting these causal structures, where wires represent quantum systems, and boxes represent operations: gates (green), state preparations (orange) and the operation of discarding a system (black). **c**, An example of a family of quantum circuits that range over the three possible causal relations. The gate acting on A and E (dashed green box) is either identity (top), swap (middle) or a probabilistic mixture of the two (bottom).

A complete solution of the causal inference problem specifies not just the causal structure but also the functional relationship that holds between each system and its causal parents. For instance, this can be achieved by specifying the identity of the gates in the circuits depicted in Fig. 1b. More generally, we aim to specify the functionality of the unknown circuit fragment that relates A and B (the dashed region in Fig. 1b).

A particular example of our causal inference problem is depicted in Fig. 1c. A system A is prepared in a maximally entangled state with an ancillary system E. Subsequently, A and E are subjected to an unknown quantum operation drawn from a one-parameter family: identity, with probability $1 - p$, and swap, with probability p. The case of pure identity corresponds to a purely cause–effect connection between A and B (top), the case of pure swap corresponds to a purely common-cause connection (middle), and every other case corresponds to a hybrid of the two causal structures (bottom).

Interventionist versus observational probing schemes

The randomized trial discussed in the introduction is an example of a classical experimental scheme that allows one to distinguish a cause–effect connection from a common-cause connection. In fact,
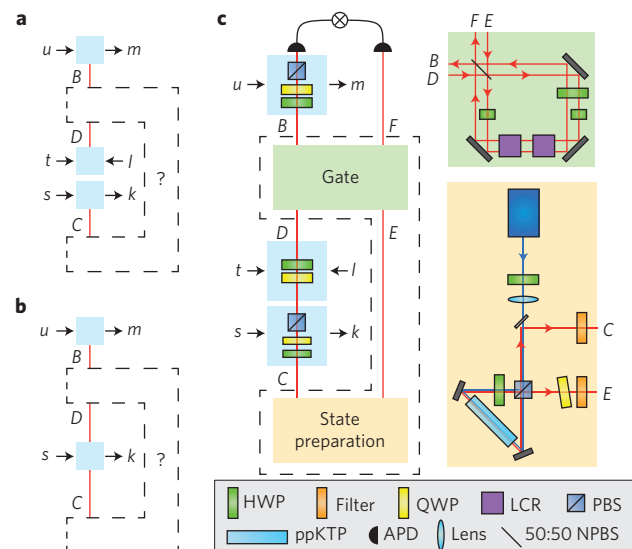
Figure 2 | Two schemes for probing causal relations and experimental set-up. The unknown circuit fragment enclosed in the dashed box can be probed by two schemes. **a**, Interventionist scheme. The outputs B and C are both subjected to tomographically complete sets of measurements, whereas D is prepared in states drawn from a tomographically complete set. Lowercase variables denote settings and outcomes of these interventions. **b**, Observational scheme. The outputs B and C are both subjected to tomographically complete sets of measurements. The measurement on C is projective, fixing the preparation on D. **c**, Experimental set-up including polarization-entangled photon source and probabilistic swap gate. Notation for optical elements: half-wave plate (HWP), quarter-wave plate (QWP), liquid crystal retarder (LCR), one of which is turned on with probability p, polarizing beamsplitter (PBS), periodically poled KTP crystal (ppKTP), avalanche photodiode (APD), and non-polarizing beamsplitter (NPBS).

there is a broad class of schemes that share this feature; we refer to them as interventionist schemes. Their defining characteristic is that the information they yield about a variable's value before the intervention is different from the information they yield about the same variable's value after the intervention. In a drug trial, for example, if the treatment is assigned to patients neither completely at random, nor exactly according to their choice, but rather is partially randomized, then one can still learn something about the effectiveness of the drug[2].

On the other hand, we define the class of observational schemes as those that yield precisely the same information about the pre-measurement state as they do about the post-measurement state, and which therefore—classically—cannot reveal the causal structure. Completely passive observation is an example of such a scheme, but there are many others. For instance, for a classical particle in one dimension, a measurement that reveals its position while randomizing its momentum is clearly not completely passive. Nonetheless, as long as the initial distribution over momentum is uniform, the probability distribution over phase space describing what is known about the particle before the measurement (conditioned on the measurement outcome) is the same as the distribution describing what is known about the particle after the measurement.

We take the property of informational symmetry to also define the distinction between observational and interventionist schemes in the quantum sphere. The quantum interventionist scheme we use mirrors a classical perfectly randomized trial: we first measure the early system (A), then reprepare it in a state selected at random. On the other hand, no quantum observational scheme can be a precise analogue of a completely passive observation because learning the

value of one quantum observable randomizes the values of other, complementary, observables. However, if the initial state of our system A is completely mixed, and if the measurement is projective and obeys the standard state update rule (wherein the system is left in the pure state in which it was found), then one obtains the same information about the pre-measurement state of A as one obtains about the post-measurement state of A. Such a scheme is consequently in the observational class.

It is useful to adopt a distinct notation for the versions of A before and after the measurement on it[6]. We denote these by C and D respectively, so that C can only be related to B by a common cause, whereas D can only have a directed causal influence on B. The two probing schemes are illustrated in Fig. 2a,b.

We will show that the quantum causal inference problem can be completely solved in the interventionist scheme, by performing informationally complete sets of measurements on B and on C and an informationally complete set of preparations on D. In the observational scheme, on the other hand, we are limited to performing an informationally complete set of measurements on B and on C, whereas the preparation of D is determined by the outcome of the measurement on C.

Each measurement is described by two classical variables: the measurement setting, specifying which observable is measured, and the outcome. We denote these by s and k respectively for the measurement on C, and by u and m for B, as depicted in Fig. 2a,b. In the interventionist scheme, system D is prepared in the l eigenstate of the t observable. Therefore, the experimental data in this case is the conditional probability distribution $P(km|lstu)$. In the observational scheme, the standard state update rule implies that D is reprepared with $t = s$ and $l = k$, so the experimental data takes the form $P(km|su)$.

Experiment

We implement the one-parameter family of circuits introduced in Fig. 1c, which ranges through the possible causal structures as we vary the parameter p, using the experimental set-up shown in Fig. 2c. The polarization degrees of freedom of a pair of photons constitute the pair of systems, which in this case are qubits. For an informationally complete set of measurements we use the three Pauli observables, hence measurement settings $\{1, 2, 3\}$ and outcomes $\{+1, -1\}$. We use downconversion to create entangled photon pairs in the state $|\Phi^+\rangle = (1/\sqrt{2})(|H\rangle|H\rangle + |V\rangle|V\rangle)$, where $|H\rangle(|V\rangle)$ denotes horizontal (vertical) polarization. One of the photons, C, is subjected to a polarization measurement, after which the photon is denoted D. In the interventionist scheme, this measurement is followed by a random repreparation of the photon's polarization. Photon D and the other photon from the downconversion pair are then subjected to a probabilistic swap gate: with probability p the modes are exchanged; otherwise they are unaffected. The first photon of the output, B, is subjected to a final measurement of its polarization before both photons are detected in coincidence.

Mathematical form of the unknown circuit fragment

The circuit fragment represented by the dashed box in Fig. 2a implements a physical process that takes one input, D, and produces outputs B and C. It can therefore be represented, as shown in the Supplementary Methods, by a completely positive and trace-preserving map of the form $\mathcal{E}_{CB|D} : \mathcal{L}(\mathcal{H}_D) \rightarrow \mathcal{L}(\mathcal{H}_C \otimes \mathcal{H}_B)$, where $\mathcal{L}(\mathcal{H})$ is the space of linear operators on the Hilbert space $\mathcal{H}$. Note, however, that the output C precedes the input D in time. The map $\mathcal{E}_{CB|D}$ must therefore satisfy the additional constraint that C cannot causally depend on D: tracing out B from $\mathcal{E}_{CB|D}$ leaves a map $\mathcal{E}_{C|D}$ that produces a fixed state ρ_C regardless of the input at D, that is, $\mathcal{E}_{C|D} = \rho_C \mathrm{Tr}_D$. We term such an object a causal map. Its properties are further discussed in the Supplementary Methods.

Circuit fragments that do not fall into one of the standard classes (preparations, channels, or measurements) have been previously considered in the context of alternative formulations of quantum theory by a number of authors, under different names: quantum combs[7], operator tensors[8], process matrices[9], quantum conditional states[10,11] and others[12,13].

Note that the causal map $\mathcal{E}_{CB|D}$ incorporates as special cases both bipartite states and unipartite processes. If the structure is purely common cause (as in Fig. 1b middle), the map will have the form

$$\mathcal{E}_{CB|D}^{cc} = \rho_{CB}\mathrm{Tr}_D \qquad (1)$$

describing a state on CB and the trace operation on D. Conversely, if the structure is purely cause–effect (as in Fig. 1b top), the map will have the form

$$\mathcal{E}_{CB|D}^{ce} = \rho_C \otimes \mathcal{E}'_{B|D} \qquad (2)$$

describing a quantum channel from D to B and a normalized state on C. To ensure that projective measurements yield purely observational schemes, we require, as noted earlier, that the initial state on C be completely mixed, that is, $\rho_C = (1/d_C)\mathbb{1}_C$ in equation (2) and $\mathrm{Tr}_B(\rho_{CB}) = (1/d_C)\mathbb{1}_C$ in equation (1), where d_C is the dimension of the Hilbert space of C.

More generally, the causal map can describe objects that are neither states nor processes (as in Fig. 1b bottom). For example, if the structure is a probabilistic mixture of common cause and cause–effect, we can write

$$\mathcal{E}_{CB|D} = p\mathcal{E}_{CB|D}^{cc} + (1-p)\mathcal{E}_{CB|D}^{ce} \qquad (3)$$

where the mixing parameter $p \in [0, 1]$ interpolates between the extreme cases given in equations (1) and (2). An even more general form arises if cause–effect and common-cause contributions act at the same time, for instance, if the example of Fig. 1c is modified to allow a family of unitaries that coherently interpolate between identity and swap.
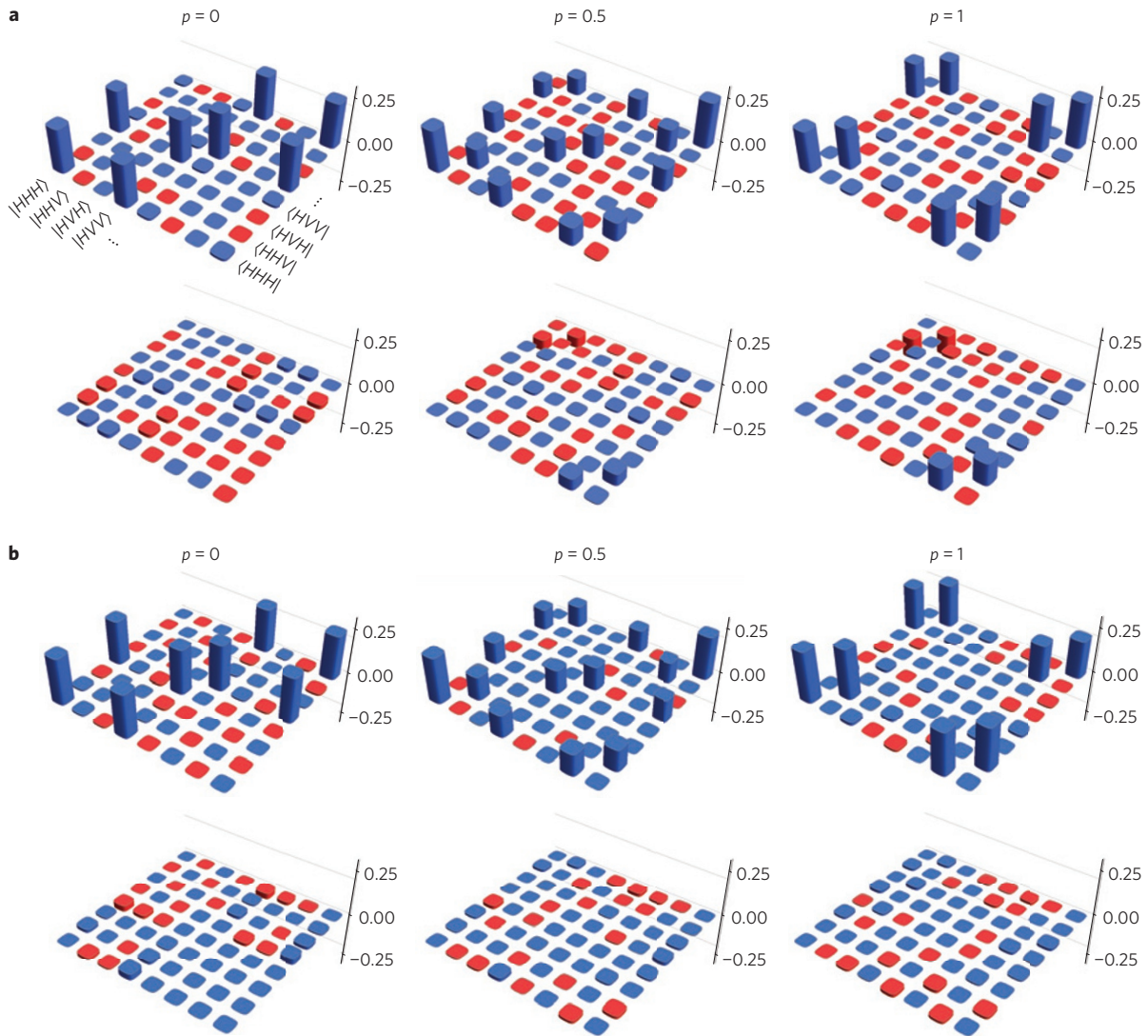
Data analysis in the interventionist scheme

The conditional probability distribution $P(km|lstu)$ obtained in the interventionist scheme is sufficient to tomographically reconstruct the map $\mathcal{E}_{CB|D}$ for systems of any dimension. This is proved in the Supplementary Methods. The key is that the set of preparations on D span $\mathcal{L}(\mathcal{H}_D)$, and the sets of measurements on B and C span $\mathcal{L}(\mathcal{H}_B)$ and $\mathcal{L}(\mathcal{H}_C)$, so that together they completely characterize the input–output functionality of the map, in the same way that informationally complete sets of preparations and measurements allow conventional tomography of states and processes. We term this scheme causal tomography as it achieves a complete solution of the causal inference problem. Considering that the map $\mathcal{E}_{CB|D}$ subsumes bipartite states and processes as special cases, but also describes more general possibilities, causal tomography constitutes a novel, more general scheme that includes conventional tomography schemes as limiting cases.

We apply our scheme to tomographically reconstruct $\mathcal{E}_{CB|D}$ from data obtained in the experiment that implements the interventionist scheme. The resulting maps are presented in Fig. 3a and are found to achieve an average fidelity of $(97.7 \pm 0.2)\%$ with our theoretical predictions. The quoted error on the fidelity reflects the effect of Poissonian noise in the observed count numbers.

Instead of a full specification of $\mathcal{E}_{CB|D}$, one is often interested in more coarse-grained information, for instance, the value of the probability p in the mixture. To estimate p in the interventionist scheme, we fit the experimental data to a map of the form of equation (3). Figure 4a shows our best estimate, p_{fit}, as a function of

Figure 3 | Reconstruction of the causal map. Reconstructions based on the interventionist scheme (**a**) and the observational scheme (**b**), for three different causal structures (shown in Fig. 1c), with probability of common cause $p=0$ (left), $p=1/2$ (middle) and $p=1$ (right). We show the Choi representation of the maps[14,15,26], which is defined as the tripartite state $\tau_{CBD} \equiv (\mathcal{E}_{CB|D'} \otimes \mathbb{1}_D)(|\Phi^+\rangle_{D'D}\langle\Phi^+|)$. The arrays represent the real (top) and imaginary (bottom) components of the density matrices, with blue representing positive values and red negative values. The basis is ordered as CBD, that is, $|HHH\rangle = |H\rangle_C|H\rangle_B|H\rangle_D$. In the case $p=0$, we expect $\mathcal{E}_{CB|D} = (1/2)\mathbb{1}_C \otimes \mathcal{I}_{B|D}$, with the identity map $\mathcal{I}_{B|D}$ from B to D, which corresponds to a Choi state $\tau_{CBD} = (1/2)\mathbb{1}_C \otimes |\Phi^+\rangle_{BD}\langle\Phi^+|$. When $p=1$, we expect $\mathcal{E}_{CB|D} = |\Phi^+\rangle_{CB}\langle\Phi^+|\mathrm{Tr}_D$, which has Choi state $\tau_{CBD} = |\Phi^+\rangle_{CB}\langle\Phi^+|\otimes(1/2)\mathbb{1}_D$. The case $p=1/2$ produces an equal mixture of the two. In the experiment that implemented the interventionist scheme, the state prepared on CE contained an inadvertent phase shift: instead of $|\Phi^+\rangle$ as planned, it is close to $(1/\sqrt{2})(|H\rangle|H\rangle + e^{i\phi}|V\rangle|V\rangle)$. Based on the correlations between C and B in the purely common-cause case, we estimated the phase shift to be $\phi = -0.17\pi$. This accounts for the non-zero imaginary components in the Choi states in the interventionist scheme. This phase shift is no impediment to causal inference; one can simply revise the Choi states expected from each causal structure to include it. The tomographically reconstructed states, denoted τ_{fit}, are compared to the expected Choi states by computing the fidelity $F \equiv (\mathrm{Tr}\sqrt{\tau^{1/2}\tau_{\mathrm{fit}}\tau^{1/2}})^2$ (ref. 16). The interventionist scheme achieves fidelities of $(98.3\pm0.1)\%$ ($p=0$), $(98.2\pm0.2)\%$ ($p=1/2$) and $(96.6\pm0.3)\%$ ($p=1$) to the expected states, which the observational scheme matches at $(98.38\pm0.07)\%$, $(98.31\pm0.07)\%$ and $(96.38\pm0.09)\%$, respectively. In light of these high fidelities, we note that the expected Choi states would appear almost indistinguishable from the reconstructed ones depicted here.

the value that the experiment sought to implement, p_{exp}. Our scheme is shown to extract p with high accuracy, with an rms deviation from the implemented value of only 0.024.

This deviation is due mostly to the way we implemented the probabilistic mixture, by dividing each measurement into 25 intervals and, for each interval, sampling from $(p_{\mathrm{exp}}, 1-p_{\mathrm{exp}})$ to decide whether identity or swap is implemented. The relative frequency of swap realized over the course of such a measurement, therefore, follows a binomial distribution about p_{exp}. The rms deviation between p_{exp} and p_{fit} due to this effect was estimated by Monte Carlo to be 0.012, comparable to the observed value.

A small systematic error in p_{fit} arises owing to the finite contrast of the swap gate: when $p_{\mathrm{exp}} = 1$, only 96% of the input photons emerge in the correct output path, whereas $p_{\mathrm{exp}} = 0$ implements identity with 99% contrast. This is reflected in Fig. 4a: as p_{exp} reaches 1.00, the best-fitting p_{fit} is only 0.96.

Data analysis in the observational scheme

Unlike the interventionist scheme, the observational scheme does not allow a complete tomographic reconstruction of the map $\mathcal{E}_{CB|D}$ in an arbitrary causal scenario. This is because, without the randomizing repreparation, the state prepared on D is the same
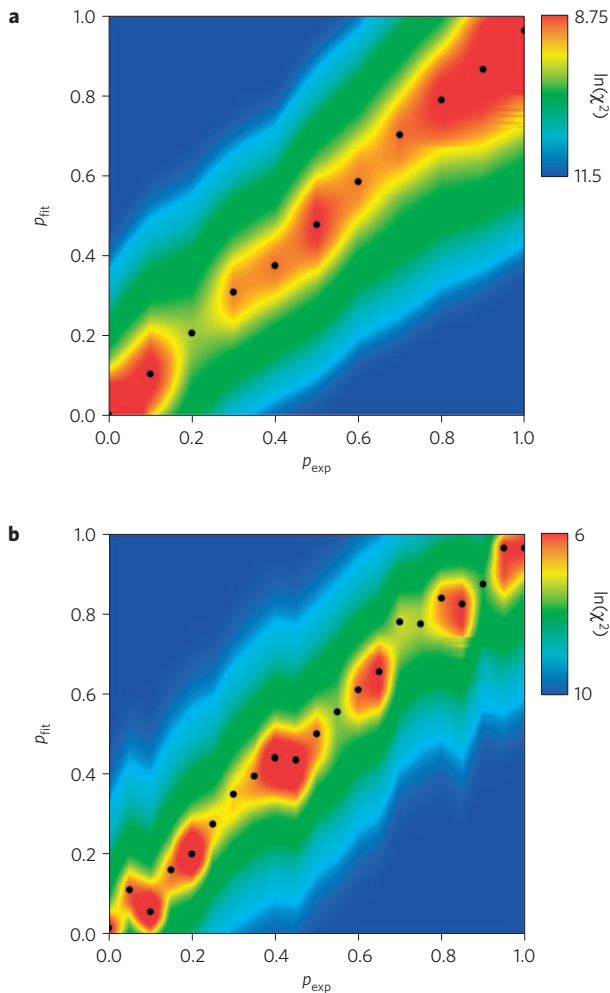
Figure 4 | Indicators of causal structure determined by interventionist and observational schemes. We probe a probabilistic mixture of common cause, with probability p_{exp}, and cause–effect, using the interventionist scheme (**a**) or the observational scheme (**b**), and fit to a mixed causal structure with probability of common cause p_{fit}. Colour encodes the quality of the fit, as measured by the logarithm of the least-squares residue χ^2 with black points marking the optimal p_{fit} for each p_{exp}. The generally large values of χ^2 are due to uncertainty in the values of p_{exp} implemented during each measurement, as discussed in the text. Despite this fact, one can clearly discern a narrow valley around $p_{fit} = p_{exp}$. This shows that our analysis recovers the correct value for the probability, thereby identifying the causal structure, with a root mean square (rms) deviation of 0.024 in the interventionist scheme and 0.032 using the observational scheme. (Fewer values of p_{exp} were sampled in the interventionist scheme because it requires more measurements.)

Table 1 | Signatures of causal structure accessible by the observational scheme.

Pattern of correlations				Cause–effect explanation?	Common-cause explanation?
C_{11}	C_{22}	C_{33}	$\Pi_v C_{vv}$		
+1	+1	+1	+1	$\mathcal{E}(\cdot) = \mathbb{1}(\cdot)\mathbb{1}$	No
+1	−1	−1	+1	$\mathcal{E}(\cdot) = \sigma_1(\cdot)\sigma_1$	No
−1	+1	−1	+1	$\mathcal{E}(\cdot) = \sigma_2(\cdot)\sigma_2$	No
−1	−1	+1	+1	$\mathcal{E}(\cdot) = \sigma_3(\cdot)\sigma_3$	No
−1	−1	−1	−1	No	$\rho = |\Psi^-\rangle\langle\Psi^-|$
−1	+1	+1	−1	No	$\rho = |\Phi^-\rangle\langle\Phi^-|$
+1	−1	+1	−1	No	$\rho = |\Phi^+\rangle\langle\Phi^+|$
+1	+1	−1	−1	No	$\rho = |\Psi^+\rangle\langle\Psi^+|$

Suppose the same Pauli observable is measured on two qubits A and B, that is, $(s, u) \in \{(1, 1), (2, 2), (3, 3)\}$, and suppose the outcomes k and m are found to be perfectly correlated, either positively or negatively, with correlation coefficients $C_{su} \equiv p(k = m|su) - p(k \neq m|su) \in \{\pm 1\}$. For simplicity, we also assume that the marginal distribution over k (respectively m) is uniform for all values s (respectively u). In this case, perfect negative correlation for all three observables can only be explained by a common cause—namely, when the measurements are implemented on two qubits prepared in the singlet state $\rho = |\Psi^-\rangle\langle\Psi^-|$. There is no channel that produces this pattern: it would constitute a universal NOT gate, which is not a completely positive map. Similarly, positive correlation for all three observables admits of a cause–effect explanation—namely, when the measurements are implemented on the input and output of the identity channel $\mathcal{E}(\cdot) = \mathbb{1}(\cdot)\mathbb{1}$. No bipartite state has this pattern of correlations, a fact sometimes described as the nonexistence of an 'antisinglet' state. Every row of the table can be explained in this fashion. The signature of the causal structure is contained, therefore, in the product of the correlation coefficients, $C_{11} \cdot C_{22} \cdot C_{33}$. Notation: σ_1, σ_2 and σ_3 are the Pauli matrices and $|\Psi^\pm\rangle \equiv (1/\sqrt{2})(|0\rangle|1\rangle \pm |1\rangle|0\rangle)$, $|\Phi^\pm\rangle \equiv 1/\sqrt{2}(|0\rangle|0\rangle \pm |1\rangle|1\rangle)$ are the Bell states, with $\{|0\rangle, |1\rangle\}$ the eigenstates of σ_3.

as the one found in the measurement on C. Therefore, although the measurements on C span the operator space $\mathcal{L}(\mathcal{H}_C)$ and the repreparations of D span $\mathcal{L}(\mathcal{H}_D)$, they do not together span the operator space $\mathcal{L}(\mathcal{H}_C \otimes \mathcal{H}_D)$.

Nonetheless, the correlations that are observed between A and B in the observational scheme may still contain a signature of the causal structure, as demonstrated by the example in Table 1. It turns out that one can perfectly distinguish any unitary process from any pure maximally entangled bipartite state. Moreover, if the causal scenario is a probabilistic mixture of these two possibilities, and the systems in question are qubits, then the observational scheme allows one to infer both the probability and the exact nature of the process and the bipartite state, thereby affording a complete solution of the

causal inference problem (up to one choice of sign). An explicit proof is provided in the Supplementary Methods.

Our experiment implements such a mixture (where the particular choice of process and state removes the sign ambiguity), and confirms that we can obtain a complete solution of the causal inference problem: Fig. 3b shows causal maps reconstructed from observational data. The average fidelity of $(97.7 \pm 0.1)\%$ is on par with the results from the interventionist scheme. Figure 4b shows the best fit for the mixing parameter p based on observational data, finding an rms deviation of 0.032 from the implemented values; only slightly worse than with the interventionist scheme. The main source of noise is again the statistical fluctuations in p_{exp} discussed before, which can account for an rms deviation of 0.043 according to a Monte Carlo simulation. Considering that causal inference about two variables based on observation alone is impossible in the classical context without additional assumptions, this result demonstrates a quantum advantage.

Discussion

The example from Table 1 makes use of entangled states and coherent channels. This is not an accident. Common-cause mechanisms that prepare separable states and cause–effect mechanisms that implement entanglement-breaking channels ('measure and reprepare') produce the same patterns of correlations under observational schemes—they do not reveal the causal structure. Entanglement and coherence are therefore necessary for achieving the quantum advantage in causal inference. If we are promised either a purely common-cause or a purely cause–effect relation between two qubits, they are also sufficient, as we show in the Supplementary Methods. For higher-dimensional systems, the advantage arises only for entangled states that have negative partial transpose and for coherent channels that are Choi-isomorphic to such states.

The causal inference schemes described here promise extensive applications in experiments exhibiting quantum effects. For instance, they can provide a test of whether the dynamics of a given open quantum system is Markovian or not[17–24]. This is because in

a non-Markovian evolution, the state of the system at some time t_2 is influenced not only directly by the state at time t_1, but also by some environmental degrees of freedom, which act as a reservoir of information about the state of the system at some earlier time t_0—the common cause. Our inference schemes may also help to detect initial correlations between system and environment, which, if unaccounted for, can lead to errors in the characterization of processes[25–30].

Our results suggest several interesting avenues for future research. What can be inferred from observational data when the state is not pure and the channel is not unitary? What is the most general sort of quantum measurement scheme that falls into the observational class and what can be inferred from it? How does one generalize such schemes from pairs of systems to arbitrary numbers of systems?

Methods

We produce polarization-entangled photons using parametric downconversion in a nonlinear crystal embedded in a Sagnac interferometer[31–33]. A 10 mW laser with centre wavelength 405 nm propagates through a polarizing beamsplitter (PBS), splitting into two components that travel in opposite directions in the Sagnac interferometer. Each component produces degenerate type-II phase-matched downconverted pairs at 809.5 nm in a 10 mm periodically poled KTP (ppKTP) crystal. There is a half-wave plate (HWP) placed on one side of the crystal so that each component enters the crystal with horizontal polarization. On exiting the interferometer through the PBS, the photon pair is entangled in polarization. The exact entangled state can be set using a HWP in the pump beam and a quarter-wave plate (QWP) in one of the exiting photon paths; we prepare the maximally entangled state $|\Phi^+\rangle$. In one arm, we separate the pump from the downconverted light using a dichroic mirror. The photons are coupled into single-mode fibres after passing through a bandpass filter to reduce background. We use quantum state tomography to characterize the source and find an average fidelity of 98.5% with $|\Phi^+\rangle$.

Our measurement set consists of horizontal $|H\rangle$ and vertical $|V\rangle$ polarization, $|D\rangle = (1/\sqrt{2})(|H\rangle + |V\rangle)$, $|A\rangle = (1/\sqrt{2})(|H\rangle - |V\rangle)$, $|R\rangle = (1/\sqrt{2})(|H\rangle + i|V\rangle)$, and $|L\rangle = (1/\sqrt{2})(|H\rangle - i|V\rangle)$. We measure polarization using a PBS preceded by a HWP and a QWP, which are adjusted so that only one particular eigenstate can pass. By alternating the settings of the wave plates to transmit either one or the other eigenstate of a given Pauli observable in different runs of the experiment, we obtain the same statistics as if we measured in a completely non-destructive manner, which would extract the eigenvalue of the desired Pauli observable while leaving the photon intact.

The experiment requires a gate $\mathcal{G}$ that can faithfully transmit the photon polarizations directly from $D \to B$ and $E \to F$ with probability $1 - p$ and swap the photon polarizations from $D \to F$ and $E \to B$ with probability p. We implement this with the displaced Sagnac interferometer shown in Fig. 2c (refs 34,35). There are two distinct paths in the interferometer: one travelling clockwise, the other travelling anticlockwise. If there is no phase difference between the two paths, the light exits the interferometer at the same side of the beamsplitter at which it entered, with a transverse displacement; if there is a π phase difference, it exits at the opposite side. If light is incident on both input ports of the interferometer, the zero phase shift implements the identity, whilst the π phase shift implements the swap.

This probabilistic switching is implemented using a variable liquid crystal retarder (LCR), whose birefringence can be controlled by an external voltage. A second LCR is included and set to perform the identity operation for compensation. Two half-wave plates at 45° are inserted into each path of the interferometer. The clockwise path encounters the LCRs after passing through both waveplates, whereas the anticlockwise path encounters the LCRs between the two waveplates. This asymmetry results in the birefringence affecting the two paths differently. When the controlled LCR implements the identity, $\mathbb{1}$, both paths pick up the same phase shift, so the net effect of the gate is the identity. When it implements the phase gate, σ_3, one path picks up a π phase shift with respect to the other, so the net effect of the gate is the swap. The LCR is controlled by the random number generator in LabView, which generates samples from the distribution $(p, 1-p)$. The switching rate of the LCR is 5 Hz and data is accumulated for 5 s for each measurement setting, yielding 25 samples for every value of the parameter p.

The experiments proceed as follows. After preparing the entangled state on modes C and E, we measure the polarization of C. Assuming that the photon passes through the PBS, we can then reprepare it with another QWP and HWP in the desired state for mode D. Light in modes D and E is then sent into the gate $\mathcal{G}$. The output of the gate in mode B is detected using a HWP, QWP and PBS, and F is directly detected. We detect in coincidence to ensure that the source

produced the requisite state. The coincidence detection is performed using single-photon detectors and coincidence logic with a window of 3 ns. Our coincidence count rate at B and F is approximately 2,000 Hz.

References

1. Reichenbach, H. *The Direction of Time* (Univ. of California Press, 1956).
2. Pearl, J. *Causality: Models, Reasoning and Inference* (Cambridge Univ. Press, 2000).
3. Spirtes, P., Glymour, C. & Scheines, R. *Causation, Prediction, and Search* (MIT Press, 2000).
4. Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J. & Schölkopf, B. Distinguishing cause from effect using observational data: Methods and benchmarks. Preprint at http://arxiv.org/abs/1412.3773 (2014).
5. Fitzsimons, J., Jones, J. & Vedral, V. Quantum correlations which imply causation. Preprint at http://arxiv.org/abs/1302.2731 (2013).
6. Richardson, T. S. & Robins, J. M. *Single World Intervention Graphs (SWIGs): A Unification of the Counterfactual and Graphical Approaches to Causality* (CSSS, University of Washington, 2013).
7. Chiribella, G., D'Ariano, G. M. & Perinotti, P. Theoretical framework for quantum networks. *Phys. Rev. A* **80,** 022339 (2009).
8. Hardy, L. The operator tensor formulation of quantum theory. *Phil. Trans. R. Soc. A* **370,** 3385–3417 (2012).
9. Oreshkov, O., Costa, F. & Brukner, C. Quantum correlations with no causal order. *Nature Commun.* **3,** 1092 (2012).
10. Leifer, M. & Spekkens, R. W. Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference. *Phys. Rev. A* **88,** 052130 (2013).
11. Leifer, M. S. Quantum dynamics as an analog of conditional probability. *Phys. Rev. A* **74,** 042310 (2006).
12. Aharonov, Y., Popescu, S., Tollaksen, J. & Vaidman, L. Multiple-time states and multiple-time measurements in quantum mechanics. *Phys. Rev. A* **79,** 052110 (2009).
13. Oeckl, R. A "general boundary" formulation for quantum mechanics and quantum gravity. *Phys. Lett. B* **575,** 318–324 (2003).
14. Choi, M. D. Completely positive linear maps on complex matrices. *Linear Algebr. Appl.* **10,** 285–290 (1975).
15. D'Ariano, G. M. & Lo Presti, P. Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation. *Phys. Rev. Lett.* **86,** 4195–4198 (2001).
16. Jozsa, R. Fidelity for mixed quantum states. *J. Mod. Opt.* **41,** 2315–2323 (1994).
17. Wolf, M. M., Eisert, J., Cubitt, T. & Cirac, J. Assessing non-Markovian dynamics. *Phys. Rev. Lett.* **101,** 150402 (2008).
18. Laine, E-M., Piilo, J. & Breuer, H-P. Measure for the non-Markovianity of quantum processes. *Phys. Rev. A* **81,** 062115 (2010).
19. Rivas, Á., Huelga, S. F. & Plenio, M. B. Quantum non-Markovianity: Characterization, quantification and detection. *Rep. Prog. Phys.* **77,** 094001 (2014).
20. Rivas, Á., Huelga, S. F. & Plenio, M. B. Entanglement and non-Markovianity of quantum evolutions. *Phys. Rev. Lett.* **105,** 050403 (2010).
21. Lu, X-M., Wang, X. & Sun, C. Quantum Fisher information flow and non-Markovian processes of open systems. *Phys. Rev. A* **82,** 042103 (2010).
22. Liu, B-H. *et al.* Experimental control of the transition from Markovian to non-Markovian dynamics of open quantum systems. *Nature Phys.* **7,** 931–934 (2011).
23. Tang, J-S. *et al.* Measuring non-Markovianity of processes with controllable system-environment interaction. *Europhys. Lett.* **97,** 10002 (2012).
24. Wallman, J., Flammia, S., Barnhill, M. & Emerson, J. Simpler, faster, better: Robust randomized benchmarking tests for non-unitality and non-Markovianity in quantum devices. *Bull. Am. Phys. Soc.* 77 (2014).
25. Pechukas, P. Reduced dynamics need not be completely positive. *Phys. Rev. Lett.* **73,** 1060–1062 (1994).
26. Altepeter, J. B. *et al.* Ancilla-assisted quantum process tomography. *Phys. Rev. Lett.* **90,** 193601 (2003).
27. Boulant, N., Emerson, J., Havel, T. F., Cory, D. G. & Furuta, S. Incoherent noise and quantum information processing. *J. Chem. Phys.* **121,** 2955–2961 (2004).
28. Weinstein, Y. S. *et al.* Quantum process tomography of the quantum Fourier transform. *J. Chem. Phys.* **121,** 6117–6133 (2004).
29. Howard, M. *et al.* Quantum process tomography and Lindblad estimation of a solid-state qubit. *New J. Phys.* **8,** 33 (2006).
30. Carteret, H., Terno, D. R. & Zyczkowski, K. Physical accessibility of non-completely positive maps. *Phys. Rev. A* **77,** 042113 (2008).
31. Kim, T., Fiorentino, M. & Wong, F. N. C. Phase-stable source of polarization-entangled photons using a polarization Sagnac interferometer. *Phys. Rev. A* **73,** 012316 (2006).

32. Fedrizzi, A., Herbst, T., Poppe, A., Jennewein, T. & Zeilinger, A. A wavelength-tunable fiber-coupled source of narrowband entangled photons. *Opt. Express* **15,** 15377–15386 (2007).
33. Biggerstaff, D. N. *et al.* Cluster-state quantum computing enhanced by high-Fidelity generalized measurements. *Phys. Rev. Lett.* **103,** 240504 (2009).
34. Kwiat, P. G., Mitchell, J. R., Schwindt, P. D. D. & White, A. G. Grover's search algorithm: An optical approach. *J. Mod. Opt.* **47,** 257–266 (2000).
35. Nagata, T., Okamoto, R., O'Brien, J. L., Sasaki, K. & Takeuchi, S. Beating the standard quantum limit with four-entangled photons. *Science* **316,** 726–729 (2007).

Author contributions

D.J. and R.W.S. conceived the original idea for the project. K.R. and R.W.S. developed the project and the theory. M.A. and K.J.R. designed the experiment. M.A. and L.V. performed the experiment. M.A., K.R. and K.J.R. performed the numerical calculations. M.A., K.R., K.J.R. and R.W.S. analysed the results. K.R., M.A. and R.W.S. wrote the first draft of the paper and all authors contributed to the final version.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to K.R.

Competing financial interests

The authors declare no competing financial interests.