How Quantum is ‘Quantum Counterfactual Communication’?

Jonte R. Hance

Quantum engineering technology laboratories, department of electrical and electronic engineering, university of Bristol, Woodland road, Bristol, BS8 1US

Quantum counterfactual communication is the recently-proposed idea of using quantum mechanics to send messages between two parties without any particles travelling between them. While this novel idea has excited massive interest, both for potential ‘un-hackable’ communications, and insight into the foundations of quantum mechanics, it has been asked whether the phenomena is truly quantum at all, or could be performed classically. We examine the idea of counterfactuality, both classically and for quantum particles, and look at the protocols proposed so far, and conclude it must, at least to strictest definitions of particle quantisation needed, be quantum.

INTRODUCTION

Recently, there has been a lot of interest in ‘quantum counterfactual communication’ [1–58]. Making use of counterfactuals as defined in physics by Penrose as “things that might have happened, although they did not in fact happen,” [29], ‘quantum counterfactual communication’ is the use of counterfactual circumstances, alongside quantum mechanics, to send information from one party to another, with no risk of the information being sent physically through the communication channel. Understandably, given the potential this raises for un-interceptable communication, this area has excited massive interest in recent years.

However, a question that has been raised is, what separates this ‘quantum counterfactual communication’ from classical counterfactual communication [23]? In this paper, we will attempt to answer this question.

To do this, firstly, we need to investigate a variety of classical counterfactual communication cases, to find the underlying structure linking them together. Secondly, a definition of counterfactuality applicable to quantum phenomena is needed, which requires considering what constitutes the ‘path’ between measurements of a quantum particle. Thirdly, a definition is needed to identify what makes a protocol ‘quantum’ - some criteria for calling something non-classical. Finally, the protocols for ‘quantum counterfactual communication’ suggested so far will be examined, and evaluated firstly by their counterfactuality, and then by their non-classicality. This will allow the evaluation of whether they meet this definition of ‘quantum counterfactual communication’, and if so, identification of what underlying structure separates it from classical counterfactual communication.

CLASSICAL COUNTERFACTUALITY

Communicating counterfactually is something long predating quantum mechanics. For instance, in the Sherlock Holmes adventure, Silver Blaze, Holmes infers that the eponymous racehorse was abducted by its own trainer, as the stable guard-dog didn’t bark. As Holmes put it, ‘the curious incident of the dog in the night-time’ was that the dog did nothing [60]. Another, more modern example, is that of Batman and the Bat-Signal. Looking up, Bruce Wayne can be certain that no crime has been committed unless the Bat-Signal appears in the sky. Therefore, the absence of the Bat-Signal counterfactually communicates to Wayne that all is well. Extending further, any inference gained by the absence of a sign (knowing the engine of a car is fine by the lack of illumination of the ‘check engine’ light, knowing no-one is calling you by the absence of a ringing phone, or even being sure that there isn’t a fire in your house by the lack of any noise from a fire alarm) can be considered a case of counterfactual communication.

Such an inference, working from something that would have happened, but didn’t, fits with Penrose’s account of counterfactuality above. However, it is complicated by the fact that only one possibility of a binary choice can be communicated in this manner. For instance, had the racehorse been kidnapped by a stranger, the dog would have barked - a case of real, rather than counterfactual communication.

This idea of counterfactual inference was elaborated upon by Maudlin. Evaluating situations where not only something would have happened but didn’t, but that are designed so, only by something happening in a certain way could it not have happened, allows us to specifically attribute causal transfer of information to the non-occurrence of an event [61]. While this is a looser link than saying that the event was caused by the non-occurrence (for instance, it can be a case of common cause for the two events), we can still use it to say, ‘if A didn’t happen, then B wouldn’t have happened’ - a counterfactual inference. While some, such as Asher Peres, say this is a theoretical nonsense, as unperformed experiments have no result [62], a cursory look at the history of science will show otherwise. We build models of possible worlds allowed by our physical laws, which are then compared to our world to counterfactually allow us to argue for the existence/lack of existence of certain features in our world. An example of this is the discovery of Nep-
tune, which occurred by considering what the planetary orbits would be if there were no additional planets than the seven known, and then observing this was not the case. Neptune’s existence was ‘counterfactually implied’ before it was directly observed.

Structure of Classical Counterfactuality

From this, we can extract the logical structure of counterfactual communication. Firstly, it is necessary that, were the inferred event to have happened, that the signalling event would also have happened. Secondly, the signalling event mustn’t happen. From this, the counterfactual inference can be made that the inferred event did not happen. Putting this into formal logic, we get

\[ A \supset B \]
\[ \neg B \]
\[ \therefore \neg A \quad (1) \]

Any scenario fitting this logical structure can be considered a case of counterfactual communication of the inferred event’s absence.

However, typically, when communicating counterfactually, we want to be certain that there is a one-to-one correspondence between the signalling event and the inferred event, in order to ensure that the absence of the inferred event can always be recognised counterfactually. To do this, we need to add in another condition: that, were the inferred event not to happen, the signalling event would not happen either (\( \neg A \supset \neg B \)). Still, as mentioned, this is not necessary for counterfactual communication.

PATH OF A QUANTUM PARTICLE

A key preparation we need to undertake before getting into the investigation of ‘quantum counterfactual communication’ is to look at the different methods of evaluating the presence, or path, of a quantum particle. This is necessary in order to allow us to establish whether the protocols we look at later are actually counterfactual.

Naïve Approach

When considering, in a quantum sense, where we can say a photon has been, our first instinct is to treat it as a spatially localised particle. However, there are problems with this - thinking back to the history of arguments for why light is viewed as being a quantum phenomenon rather than being discrete corpuscles (e.g. Young’s Two-Slit Experiment [63]), we see why this approach is unsatisfying. Therefore, a stronger criterion is needed to determine where a photon has been, rather than trying to trace a single classical trajectory.

Full Quantum Description Approach

The first non-classical approach to a particle’s location we look at is that where we examine the entire quantum mechanical description of the system. This could be done by looking at the wavefunction of the photon, but the density matrix seems more useful, given it provides a more intuitive view, and shows mixed elements of a state. This provides all the information that exists about a quantum-mechanical phenomenon at a given moment insofar as its span over all relevant modes in a system. Therefore, by associating these modes with physical positions, we should be able to see across which modes the photon is spread.

However, the issue with this is that some of these density matrix elements correspond to paths lost when the wavefunction is collapsed at the end of the process. In this case, these correspond to the photon not returning to Alice, and so information not being sent. This means, we need some way to sort these possible paths into ones that could contribute to information being sent (paths we can say the photon was on when counterfactual communication occurred) and those that don’t. This constitutes a form of post-selection (selection of state elements based on the result they could have lead to), rather than just the pre-selection given by looking at the initial state. Therefore, we need a way to factor this post-selection into our analysis.

Consistent Histories

To try and resolve this, we could use Griffith’s Consistent Histories approach [64]. In this approach, all alternate possibilities for a system are arranged into a number of histories, whereby each history is made up of a time-ordered set of events. This ‘family’ of histories forms a set of possible pathways the system can explore. Two of these histories are ‘consistent’ if they both represent possible paths from an initial to a final state allowed under the constraints of the situation. The histories are inconsistent if they don’t. For instance, this is the case when two paths evolve to two different end-points.

Applying this to our situation, we can say that a particle hasn’t gone between Bob and Alice when there isn’t a history that travels via Bob to deliver the information to Alice, within the family going from the source to Alice. Therefore, assessment of all possible histories is needed to work out consistent families, as Griffiths does for a number of protocols [56]. This analysis boils down to the idea that, regardless of how unlikely it is, if a path can be drawn across to Bob’s side and back to Alice’s detec-
tor, then, by Consistent Histories, it cannot be said to be counterfactual.

Weak Trace

The next approach considered is the weak trace, proposed by Aharonov and Albert [65]. This approach was developed as a way to examine the state between two measurements without collapsing the state.

To look at the weak trace, we first need to define weak measurements. Weak measurements are those which provide very little information for an individual run of the experiment, but because of this doesn’t disturb the system much. This is in contrast to strong (Von Neumann) measurements, which causes the system to collapse into an eigenstate of the measured operator. The weak measurement is typically taken by weakly coupling the system to a measuring device, so while little information is gathered over one run of the system, over many runs the probability distribution is obtained, allowing information to be collected that would be lost were the system strongly measured [60]. Mathematically, this is calculated by taking the expectation value of the evolution operator over the initial state, and can be physically interpreted as drawing all possible forward-evolving paths the system can take from that initial state.

However, rather than working forwards to build the evolution of the state, we want to work backwards from a given result, to investigate the paths through which the system may have evolved. For this, rather than taking the expectation value of the initial vector over the unitary that represents the evolution, we need to post-select the state we want to obtain. As an approximation, this can be imagined as tracing the initial vector forward in time, and the final vector backwards in time, and seeing where they overlap. We do this by taking the weighting, or weak value for each path operator, \( O \), as

\[
O_w = \frac{\langle \psi_f | \hat{O} | \psi_i \rangle}{\langle \psi_f | \psi_i \rangle}
\]

where \( |\psi_i\rangle \) is the initial state and \( |\psi_f\rangle \) the final state of the photon. This specific first-order approximation of the weak value is called the Two-State Vector Formalism (TSVF), as it relies on the vectors from/to two states - that at the beginning and that at the end of the protocol. This grants both the pre- and post-selection necessary, and so gives the weak value. The weak trace approach takes this weak value and views it as the criterion for photon presence. This means, if a given path-operator returns a non-zero TSVF value, there is, to first-order, a weak trace at the path it describes. More intuitively, if you draw the paths that a quantum particle could travel from its starting position, as per its evolution, and the paths that it could have come from to get to its end position, then where they overlap there is a weak trace. In our setting, as described by Vaidman, this means not being able to say the photon wasn’t there [67].

However, this brings up unintuitive results. While, in the Consistent Histories approach, a path needs to link the initial and final points, here, it only requires the paths from the initial and final states overlap at a certain point, rather than for their entire length. This means you can feasibly get a weak trace across a path on Bob’s side of the transmission channel, without a trace across the channel itself. This was demonstrated experimentally using weak measurements in nested Mach-Zehnder Interferometers (MZIs) by Danan et al [68]. If one accepts the weak trace as a valid indicator of the photon’s path, this leads to weird occurrences - such as photons seemingly ‘jumping’ discontinuously between locations [69]. This has been a key reason put forward for doubting the formalism in favour of more conventional criteria [70][72]. However, given the argument above, this approach of using a single, ‘common-sense’ trajectory for the photon doesn’t seem a valid rationale for not using weak measurement [73].

A more persuasive counterargument given is that the TSVF ignores the non-first order weak trace that a photon leaves, and so doesn’t give the entirety of the photon’s path. Vaidman admits this to some degree, insofar as saying the TSVF gives only first-order elements of the weak trace, not higher-orders. Alongside this, further analysis of Danan et al’s data shows smaller, second-order peaks not visible in their original presentation [74].
Vaidman explains this by saying that the non-local trace on any photon is also of second-order, and so this trace applies in any set-up, even if there is physical separation between objects. Further, given there are no non-local interactions in nature, this non-local weak trace cannot be strong enough to mediate any effects, so neither can a local second-order weak trace [67]. Despite this, Vaidman still cites this ‘secondary’ trace as showing a weak trace exists in Salih et al’s most recent protocol [57].

**QUANTUM AS NON-CLASSICAL**

The next thing we need to do is determine a criterion for something being quantum, to use later to evaluate the ‘quantum-ness’ of the protocols.

The first way to attempt this is to consider the differences between classical and quantum physics. For optics, classical physics is often considered to consist of everything up to and including Maxwell’s equations. These formulate light, as a type of electromagnetic phenomena, via the evolution of waves through the electromagnetic field - in this sense, ‘classical’ light is continuous, wave-based, and has no restrictions on what energies or momenta it can have [75].

Contrary to this is ‘quantised’ light, as discovered by Einstein through the photoelectric effect [70]. In this formulation, light was discovered to have some discreteness - it had a minimum unit, the photon, and allowed energy levels dependent on a universal constant, \( \hbar \). However, this is not to say light is entirely corpuscular, as thought before Young [63], as interference effects still exist between rays. Instead, light is somewhere between the two, with both wave-like and particle-like properties.

From this, it seems something counts as quantum if it relies on properties of light that only emerge at this limit, where light begins to have some particle-like features. This occurs when single photon emission and detection are involved, and so the ability of light intensity to be split across different paths is nullified through this quantisation.

However, other definitions for quantum-ness do exist. One of the most prominent of these competing definitions is that a system is quantum when it obeys quantum informational rules for probabilities, where probability amplitudes, rather than just distributions, can affect the system. These amplitudes can be positive, negative, or even complex, so long as they are normalised across the system, and can exist in linear superposition, allowing probability interference in a way that seems classically nonsensical. However, on the whole, these effects resemble those caused by wave-like interference, meaning this seems more a change in methodology than physical circumstance. The main case in which this presents factors that cannot be reconstructed classically is entanglement.

A deeply quantum phenomenon, entanglement is where the properties of two spacelike-separated objects are so intrinsically linked that, while provably indeterminate until one of the pair is measured, as soon as the measurement on one is performed, information is gained on the other [77]. Measuring one appears to non-locally collapse the other. This observable phenomena of entanglement is one of the weirdest, and most quantum, of all effects ascribed to quantum mechanics. Sadly, however, as we only deal with single photons in counterfactual communication, this ‘gold standard’ of ‘quantum-ness’ is denied to us. This means we have to take the initial definition of ‘quantum-ness’ being when we move beyond classical, wave-based electrodynamics to systems with wavelike and particle-like properties (e.g. discreteness). Ergo, we will set our criterion for a protocol being quantum as being if its structure cannot be replicated by a protocol using only classical physics. Now we have criteria for both counterfactuality and ‘quantum-ness’, we can move on to looking at the protocols proposed to be ‘quantum counterfactual communication’.

**‘QUANTUM COUNTERFACTUAL COMMUNICATION’ PROPOSALS**

Now, we move onto going through the various protocols for ‘quantum counterfactual communication’ that have been proposed. This will allow common features to be observed, and initial evaluation to be undertaken.

Ever since the idea of practical manipulation of quantum counterfactuality was first raised by Elitzur and Vaidman [78], there has been a race to exploit it for communication. An initial advance was developed by Kwiat et al, to allow highly-probable interaction-free measurement by repeated interrogation - allowing the loss to be made effectively nil in an Elitzur-Vaidman-style set-up [79]. Despite this, all protocols derived since then, until last year, have fallen into three broad categories: protocols where the communication is counterfactual only for one bit; protocols where photons travel between Alice and Bob, but in the opposite direction to the information passed between them; and protocols where photons don’t pass between Alice and Bob when information flows, but the error/loss rates for the protocol vary depending on which bit-value Bob is sending. For each of these, key protocols of the category will be demonstrated, and arguments for and against them being counterfactual will be examined.

**Elitzur-Vaidman Bomb Tester**

All counterfactual communication protocols using quantum mechanics stem from the ‘Bomb-Tester’ proposed by Elitzur and Vaidman [78]. In this thought-experiment (demonstrated in Fig. 2), a balanced MZI
FIG. 2. The experimental layout for the Elitzur-Vaidman Bomb Tester. A photon is emitted from the source (bottom-left), enters the 50:50 beam-splitter, and is spread across both paths equally. If the bomb doesn’t (and wouldn’t) explode, the photon recombines at the second (top-left) beam-splitter, and enters detector 1 with 100% probability. If the bomb does work, and is activated, the entire apparatus is destroyed. If the bomb would work, but isn’t detected at the bomb site (B), the photon has a 50:50 chance of being detected at each detector.

has one of its paths blocked by a potentially faulty bomb, that could only be ignited by a non-demolishing single-photon detection in that path.

If the photon goes down the bomb’s side of the MZI, and the bomb works, then it detonates, and the photon (and everything else) is destroyed. If the bomb doesn’t work, the photon travels through to the merging beamsplitter. However, more interestingly, if the photon doesn’t travel down the bomb’s side, there is a difference in the interference pattern created at the merging beamsplitter, and so the relative probabilities of the photon appearing at each side. This effectively allows us to test if the bomb ‘would have’ worked, without detonating it, by looking at the path the photon could have, but did not, travel down.

This differs from the classical counterfactual cases above insofar as both of the possibilities being transmitted could be said to be counterfactual. Here, this means that a valid path can be drawn from the source to the detectors that doesn’t go via the object (bomb) under detection. However, unlike the classical case, for these ‘counterfactual’ options, they are not necessarily always counterfactual. This is as, while there is the possibility for the semi-local photon to carry the information without going via the object under detection, it is not the case that the photon has to travel to the detectors without going via the detected object. This means, by all the criteria above bar the initial, naïve one, the Bomb-Tester isn’t a case of fully counterfactual communication.

Counterfactual only for One Bit

The next form of ‘quantum counterfactual communication’ considered is that where, for one bit-choice, the protocol is counterfactual, but for the other, the photon is sent to, and returns from, Bob.

Noh

FIG. 3. Noh’s protocol for counterfactual cryptography - Alice randomly polarises a photon, which passes through a beam-splitter, with one of the outputs going to Bob. There, if the photon is oppositely polarised to his guess, it passes through, is reflected back, and interferes in Alice’s interferometer, going to her \(D_2\). However, if the photon polarisation matches Bob’s guess, the wave is sent into his detector. If this clicks, the protocol is aborted, and restarted; if not, this removes the interference, and forces the photon on Alice’s side into her \(D_1\) [1].

The first of this type of protocol, and indeed the first formal ‘quantum counterfactual communication’ protocol proposed was Noh’s, published in 2009 [1]. In this set-up, he posits a more complicated scheme than those developed later. This is shown in Fig. 3. For matched polarisations, if Alice gets a click, the photon has remained on her side. However, for opposing polarisations, it has both been to, and returned from, Bob. This hardly seems counterfactual.

Despite this, the work has generated a lot of interest, with security protocols and analyses based on it still being generated [2,21]. While plenty of these have focused on reducing loss by reducing the proportion of the photon sent to Bob [2], this must always be non-zero for the protocol to function, and so there will always be some non-counterfactuality in the system. Despite this, evaluation of the bit for which the protocol doesn’t go from

Note1 Barring Guo’s adaption of the E-V Bomb Detector to use two opaque objects, which, while counterfactual insofar as the blockers, involves photons travelling between Alice and Bob in both cases [80].
Alice to Bob and back still provided a useful tool for development [22].

**Information and Photon Travel in Opposite Directions**

The next category to be investigated is that where the photon can cross the channel from Alice to Bob, but does so in the opposite direction to the information being sent. In such a set-up, there is one bit-position that causes the photon to destructively interfere across the quantum channel, keeping it at Alice, and one bit-position that causes it to constructively interfere, allowing it through to Bob. Based, therefore, on whether or not she detects it, Alice can interpret which bit was sent by Bob.

**Arvidsson-Shukur et al**

The only protocol using such a set-up is that of Arvidsson-Shukur et al. They proposed a device formed of chained MZIs, which force the photon to go to Bob if he doesn’t block, and use the Quantum Zeno effect to, for many MZIs, keep the photon at Alice if Bob does block [23]. Such a protocol is heavily derived from Kwiat et al’s Interaction Free Measurement protocol [29]. Ignoring the high probability of Alice erroneously believing that Bob hadn’t blocked (caused by the necessarily finite number of MZIs causing some probability of the photon being absorbed by a blocker), there is still the issue of the photon travelling simultaneously to the information. Considering wave-like light for a moment, the ability to carry information in the direction counter to propagation is a well-known classical phenomena, and wouldn’t be called quantum [57]. Therefore, such a protocol only seems to be ‘quantum counterfactual communication’ when you consider light as being localised, in which case, for one possible option, the photon travels from Alice to Bob. However, this means there is obviously a weak trace at Bob, and so renders it un-counterfactual for that bit-value. Therefore, while Arvidsson-Shukur et al attempt to advance arguments in favour of their protocol from the grounds of error-tolerance [25], or by calling other protocols classical through a model of classicality allowing extra, non-trivial resources, such as a shared clock, between Alice and Bob [26], they fail to give us any reason to view their protocol as true ‘quantum counterfactual communication’, and therefore their work with Calafell et al [27] merely demonstrates contra-directional information flow, an accepted classical principle.

**Photon only Travels Erroneously - Unequal Losses**

The first protocol claiming to allow fully counterfactual communication for both values of a bit sent from Bob to Alice was that of Salih et al in 2013 [28, 29]. This protocol is formed of a chain of outer interferometers, each containing a chain of inner interferometers. Similarly to the chained interferometers above, when a photon enters the inner chain, one of two things happens. If Bob blocks, the photon stays on Alice’s side (in the limit of an infinite chain, otherwise there is a chance of being absorbed by Bob’s blockers). If, however, Bob doesn’t block, and the photon were in the inner interferometer, the photon is sent to a loss channel. To that end, the outer interferometer splits some of the probability of the photon, and so has a chance of having the photon appear out from the outer arm, giving a method of Alice telling if Bob didn’t block from a method other than implication. This splitting, similarly to that in the inner interferometers, is modulated by polarisation, with small polarisation rotators before the entrance to each interferometer. Further, the outer interferometers are chained together, again going to infinity, so as to reduce losses from each outer cycle, due to not blocking, to nil.

However, as pointed out by Vaidman, there are a number of issues with the counterfactuality of the protocol, noticed when evaluating this protocol by the Weak Trace criterion (see Fig. 1) [67]. He observes a weak trace on Bob’s side of the channel through the TSVF when Bob doesn’t block, indicating that it can’t be said that the photon wasn’t there [36, 58]. While Salih et al attempt to use the commonsensical argument that, were the photon to go into the inner interferometer, it would be lost to the detector on Bob’s side [37], Vaidman’s argument held for the weak value analysis.
There are also issues raised by Griffiths from his Consistent Histories criterion, whereby he claimed a history could be traced to Bob’s side and back when Bob doesn’t block [50, 51]. However, this was countered by Salih et al later, as Griffiths only considered physical paths available, rather than the polarisations, which provide an extra degree of modal separation [47].

To fix the issue raised by Vaidman, Salih et al brought out an adjusted protocol in 2018 [48]. This protocol avoids having a weak trace on Bob’s side by using polarisation to make it so that the only waves that go to Bob are H-polarised, which are lost via $D_3$ on Bob’s side, restarting the protocol. In such a way, when Bob doesn’t block, the nature of the polarisation-based beam-splitters means there is no overlap between the forward- and backward-travelling states on Bob’s side - indicating, by the Weak Trace argument, that the photon didn’t have any presence there. This was tested practically through weak measurement, using the method proposed by Danan et al [68] and shown to have no weak trace in practice from Bob’s side visible from Alice’s detectors. Also, the full family of histories available to the photon was assessed, and once again no history was found that includes both the photon being at Bob and arriving at one of Alice’s detectors.

As was the case with Noh’s protocol, as soon as Salih et al’s protocol was published, various practical manifestations began to appear [18, 30–35, 37–46, 50–52, 81]. While these will need to be augmented to adjust for the difference between the 2013 and 2018 protocol, implementations based on the latter have already started to appear, alongside wider modal-style analysis of the protocol [39, 53].

Aharonov & Vaidman

Despite originally claiming counterfactual communication for both bit-values was impossible, at roughly the same time as Salih et al’s alteration, Vaidman, alongside Aharonov, also released a protocol to allow the communication for both bits [54]. This method is effectively the same chained set of inner interferometers as in Salih’s protocol - however, in order to avoid a weak trace in such a set-up, where there isn’t a polarisation degree of separation, at least two inner interferometers are needed. Alongside this, Aharonov and Vaidman make repeated reference to a double-sided mirror in their protocol - although, all this has the effect of doing is connecting the two inner interferometers, and folding the outer path to reduce physical space used, and so is broadly irrelevant to the counterfactuality of their protocol.

Zhang et al - Probabilistic

Zhang et al introduced a protocol, based on Salih et al’s original method, to communicate counterfactually probabilistically. In such a protocol, by their own admission, there is no guarantee of counterfactuality, but they claim the chance of it occurring is reducible to practically zero, and losses due to noise and blocking are heavily reduced [55]. However, their assertion that this protocol is counterfactual most of the time is based on a naïve local understanding of counterfactuality (whereby the photon only traces one path) rather than Consistent Histories or Weak Values, and so fails these tests as to counterfactuality.

Photon only Travels Erroneously - Equal Errors

Vaidman

Shortly after Salih et al brought out their updated protocol, Vaidman attempted to create another version of a weak trace-free counterfactual communication protocol. However, this protocol avoids the risk of an erroneous reading that Salih et al’s, and his and Aharonov’s, protocols bring [57]. Once again, it is based on a chained MZI set-up, but, as an augmentation of Salih et al’s protocol, uses the interference from the photon passing through the inner interferometer when Bob blocks to alter which detector the photon ends up at. This allows Alice, when she receives a bit, to be certain that it is the same one Bob sent. Once again, like his protocol with Aharonov, it requires more than one inner interferometer to be fully counterfactual, as otherwise the lack of the degree of freedom afforded by polarisation means a weak trace can be shown on Bob’s side using the TSVF. However, another benefit the protocol brought was that, at least for certain values of splitting in the beamsplitters in the two inner interferometer case, losses were the same for Bob blocking and not blocking.

This means, as described, no inference can be made about whether Bob blocked or not, based simply on whether a photon arrived at Alice or not. Through this, it isn’t reducible into a protocol where the information and photon are travelling simultaneously, in opposite directions. This means, by both the TSVF and the Consistent Histories approach, the counterfactuality of this protocol is indubitable.

Protocol Evaluation

By the conclusions above, something is defined as quantum if it cannot be replicated by classical physics. Here that means that using such a protocol, the same
results could not be obtained by treating light as a continuous wave. To evaluate this, we imagine what would happen if we introduced coherent light into the in-port of one of the ‘quantum counterfactual communication’ protocols.

In this scenario, the first thing we notice is, with a beam-splitter, where, when photons are used, the probability of it going in either direction is split, here the beam intensity is split. Therefore, naively, we would say that all the paths in the system are explored, and so light must at the least cross from Alice to Bob. However, as the light still interferes in the same way it would in quantum optics, when Bob doesn’t block, the waves on Alice and Bob’s side negatively interfere, so the light never returns to Alice. This means, by the weakest definition at least, the protocol remains counterfactual. This case has broadly been described previously by Gisin in his protocol for counterfactual communication, where he describes being able to perform it classically, using a Salih-style setup [23]. Such a capability is demonstrated by the fact that, in attempting to improve efficiency of Noh’s protocol, Yin et al showed it could be undertaken using a weak coherent source - accidentally proving the classicality of such a protocol [5].

Whether the classical-light-based protocol obeys stronger definitions of counterfactuality, however, is up for debate. Given the splitting nature, light will always be at both Bob and Alice. The only way to avoid this seems to be to force some locality upon the light, which can only be done by using photons, as they are forced to be spatially fixed to a single path. Therefore, the only way to make the protocol ‘strongly’ counterfactual is to use photons, and so to make the protocol quantum.

This seems somewhat similar to the case of the Elitzur-Vaidman Bomb Detector [78]. In such a setup, the existence of interference effects is entirely rational and classical for as long as classical light is being passed through the apparatus. This is as light goes to both sides of the interferometer, and the detection, and subsequent explosion, of the bomb is expected. However, when a photon is applied, and forcibly localised to one path or the other, interference becomes far spookier, and the bomb can be detected without any measurable light passing on its path. In much the same way, considered classically, information passing contra-directionally to a wave is entirely rational, and can be observed in plenty of other waves than light alone. However, when quantised, detection can occur despite the photon not having touched Bob’s side. This makes it again become something quantum, irreproducible by classical effects, and worth investigating further. This idea was brought up by Shenoy et al briefly, when considering how the existence of an Eve in the channel would affect the sending of information counterfactually [13].

CONCLUSION

From this discussion, we can clearly state that ‘quantum counterfactual communication’ is quantum. This is as, much like the Elitzur-Vaidman Bomb Tester before it, its more peculiar effects only begin to appear when a quantum particle is introduced into the system; quantum-ness is necessary in order to allow the sending of both bit-values counterfactually, rather than a single counterfactually-definite option [52]. In all schemes proposed so far, this is the only manner in which it is quantum - barring a few tentative, unproven papers on the topic, even the information being sent is classical (bits, rather than qubits). However, that does not mean it is uninteresting. ‘Quantum counterfactual communication’ allows us to look at principles at the heart of the foundations of quantum physics - self-interference and counterfactual non-definiteness - in a new and exciting way, and will hopefully further motivate thought experiments based around this seemingly nonsensical phenomenon.

Acknowledgements - I would like to thank James Ladyman, Hatim Salih, Will McCutcheon, Paul Skrzypczyk and John Rarity for useful discussions and support. I would also like to thank the Quantum Engineering Technology Laboratories at the University of Bristol, which are currently hosting me. This work was indirectly supported by the UK’s Engineering and Physical Sciences Research Council (Grants EP/T001011/1 and EP/L024020/1).

[1] Tae-Gon Noh, Counterfactual quantum cryptography. Phys. Rev. Lett., 103:230501, Dec 2009.
[2] Ying Sun and Qiao-Yan Wen. Counterfactual quantum key distribution with high efficiency. Physical Review A, 82(5):052318, 2010.
[3] Zhen-Qiang Yin, Hong-Wei Li, Wei Chen, Zheng-Fu Han, and Guang-Can Guo. Security of counterfactual quantum cryptography. Physical Review A, 82(4):042335, 2010.
[4] Musheng Jiang, Shihai Sun, and Linmei Liang. Practical stabilization of counterfactual quantum cryptography. Journal of Quantum Information Science, 1(03):116, 2011.
[5] M. Ren, G. Wu, E. Wu, and H. Zeng. Experimental demonstration of counterfactual quantum key distribution. Laser Physics, 21(4):755–760, Apr 2011.
[6] Giorgio Brida, Andrea Cavanna, Ivo Pietro Degiovanni, Marco Genovese, and Paolo Traina. Experimental realization of counterfactual quantum cryptography. Laser Physics Letters, 9(3):247, 2012.
[7] Yang Liu, Lei Ju, Xiao-Lei Liang, Shi-Biao Tang, Guo-Liang Shen Tu, Lei Zhou, Cheng-Zhi Peng, Kai Chen, Teng-Yun Chen, Zeng-Bing Chen, and Jian-Wei Pan. Experimental demonstration of counterfactual quantum communication. Phys. Rev. Lett., 109:030501, Jul 2012.

* jonte.hance@bristol.ac.uk
[8] Zhen-Qiang Yin, Hong-Wei Li, Yao Yao, Chun-Mei Zhang, Shuang Wang, Wei Chen, Guang-Can Guo, and Zheng-Fu Han. Counterfactual quantum cryptography based on weak coherent states. Physical Review A, 86(2):022313, 2012.

[9] Sheng Zhang, Jian Wang, and Chao-Jing Tang. Counterfactual attack on counterfactual quantum key distribution. EPL (Europhysics Letters), 98(3):30012, May 2012.

[10] Sheng Zhang, Jian Wang, and Chao-Jing Tang. Security proof of counterfactual quantum cryptography against general intercept-resend attacks and its vulnerability. Chinese Physics B, 21(6):060303, Jun 2012.

[11] Yan-Bing Li, Qiao-Yan Wen, and Zi-Chen Li. Security flaw of counterfactual quantum cryptography in practical setting. arXiv preprint arXiv:1312.1436, 2013.

[12] Akshata Shenoy, R. Srikanth, and T. Srinivas. Counterfactual quantum key distribution without polarization encoding. AQIS 2013 Proceedings, 2013.

[13] Akshata Shenoy-Hejamadi, R Srikanth, and T. Srinivas. Semi-counterfactual cryptography. EPL (Europhysics Letters), 103(6):60008, Sep 2013.

[14] Jia-Li Zhang, Fen-Zhuo Guo, Fei Gao, Bin Liu, and Qiao-Yan Wen. Private database queries based on counterfactual quantum key distribution. Phys. Rev. A, 88:022334, Aug 2013.

[15] Yan-Bing Li. Analysis of counterfactual quantum key distribution using error-correcting theory. Quantum information processing, 13(10):2325–2342, 2014.

[16] Xingtong Liu, Bo Zhang, Jian Wang, Chaojing Tang, Jingjing Zhao, and Sheng Zhang. Eavesdropping on counterfactual quantum key distribution with finite resources. Physical Review A, 90(2):022318, 2014.

[17] Akshata Shenoy-Hejamadi., R. Srikanth, and T. Srinivas. Counterfactual quantum certificate authorization. Phys. Rev. A, 89:052307, May 2014.

[18] Akshata Shenoy-Hejamadi and R. Srikanth. Counterfactual distribution of schrödinger cat states. Phys. Rev. A, 92:062308, Dec 2015.

[19] Tian-Yin Wang, Yan-Ping Li, and Rui-Ling Zhang. Analysis of counterfactual quantum certificate authorization. International Journal of Theoretical Physics, 55(12):5331–5335, 2016.

[20] Xiuxing Yang, Kejin Wei, Haiqiang Ma, Shihai Sun, Yungang Du, and Lingan Wu. Trojan horse attacks on counterfactual quantum key distribution. Physics Letters A, 380(18-19):1589–1592, 2016.

[21] Ya-Qi Song and Li Yang. Quantum bit commitment. Physical Review A, 99:060102, Jun 2019.

[22] I Alonso Calafell, T Strömberg, DRM Arvidsson-Shukur, LA Rozema, V Saggio, C Greganti, NC Harris, M Prabhu, J Carolan, M Hoehberg, T. Baehr-Jones, D. Englund, C.H.W. Barnes, and P. Walther. Trace-free counterfactual communication with a nanophotonic processor. npj Quantum Information, 5(1):61, 2019.

[23] Hatim Salah, Zheng-Hong Li, Mohammad Al-Amri, and Muhammad Suhail Zubairy. Protocol for direct counterfactual quantum communication. Phys. Rev. Lett., 110:170502, Apr 2013.

[24] Hatim Salah, Zheng-Hong Li, Mohammad Al-Amri, and Muhammad Suhail Zubairy. Optical communication with invisible photons. In The Rochester Conferences on Coherence and Quantum Optics and the Quantum Information and Measurement meeting, page T4B.1. Optical Society of America, 2013.

[25] Yuan Cao, Yu-Huai Li, Zhu Cao, Yuan Yin, Yuao Chen, Xiongfeng Ma, Cheng-Zhi Peng, and Jian-Wei Pan. Direct counterfactual communication with single photons. In CLEO: QELS,Fundamental Science, pages FMA4A–6. Optical Society of America, 2014.

[26] Qi Guo, Liu-Yong Cheng, Li Chen, Hong-Fu Wang, and Shou Zhang. Counterfactual entanglement distribution without transmitting any particles. Optics express, 22(8):8970–8984, 2014.

[27] Qi Guo, Liu-Yong Cheng, Li Chen, Hong-Fu Wang, and Shou Zhang. Counterfactual distributed controlled-phase gate for quantum-dot spin qubits in double-sided optical microcavities. Physical Review A, 90(4):042327, 2014.

[28] Hatim Salah. Tripartite counterfactual quantum cryptography. Physical Review A, 90(1):012333, 2014.

[29] Sheng Zhang, Bo Zhang, and Xing-tong Liu. Improved direct counterfactual quantum communication. arXiv preprint arXiv:1410.2769, 2014.

[30] Chitra Shukla and Anirban Pathak. Orthogonal-state-based deterministic secure quantum communication without actual transmission of the message qubits. Quantum Information Processing, 13(9):2099–2113, Sep 2014.

[31] Lev Vaidman. Comment on “protocol for direct counterfactual quantum communication”. Phys. Rev. Lett., 112:208901, May 2014.

[32] Hatim Salah, Zheng-Hong Li, Mohammad Al-Amri, and Muhammad Suhail Zubairy. Salah et al. reply:. Phys. Rev. Lett., 112:208902, May 2014.

[33] Muhammad Suhail Zubairy, Zhenghong Li, Mohammad D. Al-Amri, and Hatim A. Salah. Method and apparatus for direct counterfactual quantum communication, Nov 2014. US Patent 8,891,767.

[34] Yuanyuan Chen, Xuemei Gu, Dong Jiang, Ling Xie, and Lijun Chen. Tripartite counterfactual entanglement distribution. Optics express, 23(16):21193–21203, 2015.

[35] Qi Guo, Liu-Yong Cheng, Li Chen, Hong-Fu Wang, and Shou Zhang. Counterfactual distributed controlled-phase gate for quantum-dot spin qubits in double-sided optical microcavities. Physical Review A, 90(4):042327, 2014.

[36] Muhammad Suhail Zubairy, Zhenghong Li, Mohammad D. Al-Amri, and Hatim A. Salah. Method and apparatus for direct counterfactual quantum communication, Nov 2014. US Patent 8,891,767.

[37] Yuanyuan Chen, Xuemei Gu, Dong Jiang, Ling Xie, and Lijun Chen. Tripartite counterfactual entanglement distribution. Optics express, 23(16):21193–21203, 2015.

[38] Qi Guo, Liu-Yong Cheng, Li Chen, Hong-Fu Wang, and Shou Zhang. Counterfactual quantum-information transfer without transmitting any physical particles. Scientific reports, 5:8416, 2015.

[39] N Al Amri, Zheng-Hong Li, and Mohammad Suhail Zubairy. Quantum teleportation without classical channel. In Second International Seminar on Photonics, Optics, and Its Applications (ISPhOA 2016), volume 10150, page 101501N. International Society for Optics and Photonics, 2016.
[42] Yuanyuan Chen, Dong Jiang, Xuemei Gu, Ling Xie, and Lijun Chen. Counterfactual entanglement distribution using quantum dot spins. *J. Opt. Soc. Am. B*, 33(4):663–669, Apr 2016.

[43] Hatim Salih. Protocol for counterfactually transporting an unknown qubit. *Frontiers in Physics*, 3:94, 2016.

[44] Yuan Cao, Yu-Huai Li, Zhu Cao, Juan Yin, Yu-Ao Chen, Hua-Lei Yin, Teng-Yun Chen, Xiongfeng Ma, Cheng-Zhi Peng, and Jian-Wei Pan. Direct counterfactual communication via quantum zeno effect. *Proceedings of the National Academy of Sciences*, 2017.

[45] Qi Guo, Shuqin Zhai, Liu-Yong Cheng, Hong-Fu Wang, and Shou Zhang. Counterfactual quantum cloning without transmitting any physical particles. *Phys. Rev. A*, 96:052335, Nov 2017.

[46] Qi Guo, Liu-Yong Cheng, Hong-Fu Wang, and Shou Zhang. Counterfactual quantum entanglement swapping enables high-efficiency entanglement distribution. *Optics express*, 26(21):27314–27325, 2018.

[47] Hatim Salih. Comment on “particle path through a nested mach-zehnder interferometer”. *Phys. Rev. A*, 97:026101, Feb 2018.

[48] Hatim Salih, Will McCutcheon, Jonte Hance, Paul Skrzypczyk, and John Rarity. Do the laws of physics prohibit counterfactual communication? *arXiv preprint arXiv:1806.01257*, 2018.

[49] Hatim Salih. From a quantum paradox to provably counterfactual disembodied transport. *arXiv preprint arXiv:1807.06586*, 2018.

[50] Hatim Salih. Counterfactual quantum erasure: spooky action without entanglement. *Royal Society open science*, 5(2):171250, 2018.

[51] Chao Liu, Jinhong Liu, Junxiang Zhang, and Shiyao Zhu. Improvement of reliability in multi-interferometer-based counterfactual deterministic communication with dissipation compensation. *Optics express*, 26(3):2261–2269, 2018.

[52] Fakhar Zaman, Youngmin Jeong, and Hyundong Shin. Counterfactual bell-state analysis. *Scientific reports*, 8(1):14641, 2018.

[53] Jonte Hance, Will McCutcheon, Patrick Yard, and John Rarity. Modal, truly counterfactual communication protocols. *arXiv preprint arXiv:1805.12171*, 2018.

[54] Uri Peleg and Lev Vaidman. Comment on” past of a quantum particle revisited” *Annals of Physics*, 388:382–389, 2018.

[55] Yuanyuan Chen, Xuemei Gu, Dong Jiang, Ling Xie, and Lijun Chen. Counterfactual quantum cryptography network with untrusted relay. *International Journal of Modern Physics B*, 29(20):1550134, 2015.

[56] Arthur Conan Doyle. *The Memoirs of Sherlock Holmes*. George Newnes, 1894.

[57] Tim Maudlin. *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics*. Blackwell, 2002.

[58] Asher Peres. Unperformed experiments have no results. *American Journal of Physics*, 46(7):745–747, 1978.

[59] Roger Penrose. *Shadows of the Mind*, volume 4. Oxford University Press Oxford, 1994.

[60] Bob Berthelson, Alain Aspect, and Jean Dalibard. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Phys. Rev. Lett.*, 60:1351–1354, Apr 1988.

[61] Boaz Tamir and Eliehu Cohen. Introduction to weak measurements and weak values. *Quanta*, 2(1):7–17, 2013.

[62] Lev Vaidman. Past of a quantum particle. *Phys. Rev. A*, 87:052104, May 2013.

[63] Ariel Danan, Dmity Farfurnik, Shimshon Bar-Ad, and Lev Vaidman. Asking photons where they have been. *Physical review letters*, 111(24):240402, 2013.

[64] Yakir Aharonov, Eliehu Cohen, Ariel Landau, and Avshalom C. Elitzur. The case of the disappearing (and re-appearing) particle. *Scientific reports*, 7(1):531, 2017.

[65] Dmitri Sokolovskii. Are the weak measurements really measurements? *Quanta*, 2(1):50–57, 2013.

[66] Dmitri Sokolovskii. Asking photons where they have been in plain language. *Physics Letters A*, 381(4):227 – 232, 2017.

[67] David J. Griffiths. *Introduction to electrodynamics*, 2005.

[68] Albert Einstein. Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt. *Annalen der Physik*, 322:132–148, 1905.

[69] John S. Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38:447–452, Jul 1966.

[70] Asher Peres. Unperformed experiments have no results. *Annals of Physics*, 29(20):1550134, 2015.

[71] Albert Einstein. Über einen die erzeugung und verwandlung des lichtes betreffenden heuristischen gesichtspunkt. *Annalen der Physik*, 322:132–148, 1905.

[72] John S. Bell. On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38:447–452, Jul 1966.

[73] Asher Peres. Probabilistic direct counterfactual quan-