Hance, J. R., Ladyman, J., & Rarity, J. (Accepted/In press). How Quantum is Quantum Counterfactual Communication? *arXiv*. https://arxiv.org/abs/1909.07530

arXiv:1909.07530v2 [quant-ph] 30 Oct 2019

# How Quantum is Quantum Counterfactual Communication?

Jonte R. Hance,[1,*] James Ladyman,[2] and John Rarity[1]
[1] *Quantum Engineering Technology Laboratories, Department of Electrical and Electronic Engineering, University of Bristol, Woodland Road, Bristol, BS8 1US, UK*
[2] *Department of Philosophy, University of Bristol, Cotham House, Bristol, BS6 6JL*

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 has excited massive interest, both for potential 'un-hackable' communication, and insight into the foundations of quantum mechanics, it has been asked whether this phenomena is truly quantum, or could be performed classically. We examine counterfactuality, both classical and quantum, and the protocols proposed so far, and conclude it must be quantum, at least insofar as it requires particle quantisation.

## I. INTRODUCTION

Quantum Counterfactual Communication is the combination of counterfactual circumstances (where "things... might have happened, although they did not in fact happen" [1]) with quantum mechanics, to send information between two parties without any particles going between them. Given its interesting foundational implications, and potential for 'un-hackable' communication, it has excited massive interest in recent years [2–64].

What, though, separates this Quantum Counterfactual Communication from classical counterfactual communication [2]? To answer this, we need to do three things: determine the underlying structure of classical counterfactual communication; define counterfactuality for quantum phenomena, which involves considering what constitutes a particle's path between measurements; and identify what makes a protocol quantum.

Once we have done these, we can examine protocols suggested so far, to evaluate their counterfactuality and non-classicality. This will allow us to see if they meet this definition of Quantum Counterfactual Communication, and if so, identify what separates it from classical counterfactual communication.

## II. CLASSICAL COUNTERFACTUALITY

Counterfactual communication long predates quantum mechanics. For instance, in the Sherlock Holmes story, *Silver Blaze*, Holmes infers a 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 [65]. A more modern example is the Bat-Signal. Looking at the sky, Bruce Wayne can infer no crime is being committed unless the Bat-Signal appears - the Bat-Signal's absence counterfactually communicates all is well. Whenever we infer information from a sign's absence, we are being communicated to counterfactually (e.g. knowing a car's engine works by the 'check engine' light being off; knowing no-one is calling you by your phone not ringing; or being sure your house is not on fire by hearing no alarm). Inferring like this, from something that would have happened, but didn't, fits our definition of counterfactuality.

However, only one option of the two can be communicated this way. For instance, had a stranger kidnapped the racehorse, the dog would have barked - real, not counterfactual, communication. Maudlin elaborated on this, saying a sign's absence can causally transmit information - but only if the sign always occurs unless the inferred event doesn't [66]. This is less strict than saying the inferred event's non-occurrence caused the the sign's absence (as they could have a common cause), but it still lets us say, if A didn't happen, then B wouldn't have happened - a counterfactual inference. While some (such as Asher Peres) say this is nonsense, as unperformed experiments have no result [67], this seems dubious. Empiricism involves modelling possible worlds based on hypotheses, and comparing with our observations, to allow us to argue something does/does not exist. An example of this is the discovery of Neptune, where Bouvard considered what the solar orbits would be were there only the seven planets then known, then observed something different. Neptune's existence was counterfactually implied before it was directly observed.

From this, we can obtain counterfactual communication's structure. First, we need it so, were the inferred event to happen, the signalling event would too. Second, the signalling event must not happen. From this, we can deduce the inferred event did not happen. Formally,

$$A \supset B;\ \neg B;\ \therefore \neg A \tag{1}$$

Any case of this structure is counterfactual communication. However, typically, we want to be certain there is a one-to-one correspondence between the signalling and inferred event, so we always recognise the inferred event's absence. For this, we need another condition: that, if the inferred event did not happen, neither would the signalling event ($\neg A \supset \neg B$).

---

* jonte.hance@bristol.ac.uk

## III. PATH OF A QUANTUM PARTICLE

Before investigating Quantum Counterfactual Communication, we need criteria for a quantum particle's presence/path. This will allow us to establish if protocols are actually counterfactual.

### A. Naïve Approach

When considering where a quantum particle has been, our first instinct is to treat it as spatially local. However, as quantum phenomena are not solely particle-like, but as also have wave-like properties (as the Two-Slit Experiment shows [68]), this is not the case. Therefore, we need a stronger criterion for determining the path.

### B. Full Quantum Description Approach

Our first non-classical approach to a particle's location is using the entire quantum mechanical description of the system. We can do this by looking at the density matrix, which shows all of a state's elements - these provide all the information that exists about a state at a given moment. By associating them with physical positions, we can see the spread of the particle's possible locations.

However, some of these elements correspond to paths lost when the wavefunction is collapsed at the protocol's end, and so information not being sent. We need some way to sort these possible paths into those where information is sent (that the particle could have been on when counterfactual communication occurred) and those where it is not. This requires post-selection (selection of elements based on the final state they lead to), rather than just pre-selection from an initial state.

### C. Consistent Histories

To resolve this, we could use the Consistent Histories approach [69]. Here, the projectors representing possible ways for a system to evolve are time-ordered and arranged into a number of histories. This family of histories contains all possible pathways a system can evolve along from a given initial, to given final, state. These histories are consistent if they are all mutually orthogonal, as we can work out their relative probabilities. If the family is not consistent, it is meaningless to ask which path a particle took, as paths don't have valid relative probabilities.

Therefore, we can say a particle has not gone between Alice and Bob when all histories where it travels between the two (where information is transmitted), have probability zero. This requires us to analyse all possible histories in this family, as Griffiths does for a number of protocols [3]. In essence, if a particle can go between Alice and Bob when information is transmitted, by Consistent Histories it is not counterfactual.

### D. Weak Trace

Next, we consider weak measurement [70], developed to examine the state between measurements without collapsing it. Weak measurement involves lightly coupling a system to a measuring device, so while little information is gathered over one run of the system, over many runs a probability distribution is obtained. This contrasts with strong (Von Neumann) measurements, which cause a system to collapse into an eigenstate of a measured operator. Weak measurement allows us to collect information that would be lost were the system strongly measured [71]. We calculate this by taking the expectation value of the evolution operator on the initial state, and can interpret it as evaluating all possible forward-evolving paths from that initial state.

However, rather than working forwards, can also work back from a given result (post-select), to investigate the paths the system may have evolved through. If we pre- and post-select like this, we say a particle leaves a weak trace (indicating possible presence) wherever this weak measurement value is non-zero.

To approximate this trace, we can trace the initial vector forward, and the final vector backward, in time, and see where they overlap. If we represent the evolution of the state along a given path by the operator $\hat{O}$, we get this approximate value as

$$O_w = \frac{\langle\psi_f|\,\hat{O}\,|\psi_i\rangle}{\langle\psi_f|\psi_i\rangle} \tag{2}$$

where $|\psi_i\rangle$ is the initial state of the particle, and $|\psi_f\rangle$ the final. This approximation of the weak value (to the first order in trace magnitude, $\mathcal{O}(\epsilon)$) is called the Two-State Vector Formalism (TSVF), as it uses the vectors from/to two states - that at the beginning of the protocol, and that at the end. This gives both pre- and post-selection needed.

If an operator returns a non-zero TSVF value, there is to $\mathcal{O}(\epsilon)$ a weak trace along the path it describes. If we trace the paths a quantum particle could evolve along from its initial, and those it could have come from to get to its final, state, there is a weak trace where they overlap - so we cannot say the particle was not there [34].

However, this has unintuitive results. While, with Consistent Histories, a path needs to link the initial and final states, here, it only requires paths from the initial and final states overlap at some point. This means Bob can have a weak trace on his side of the transmission channel, without any in the channel itself. This was demonstrated using weak measurements in nested Mach-Zehnder Interferometers (MZIs) by Danan et al [72].

If one accepts the weak trace as a valid indicator of a

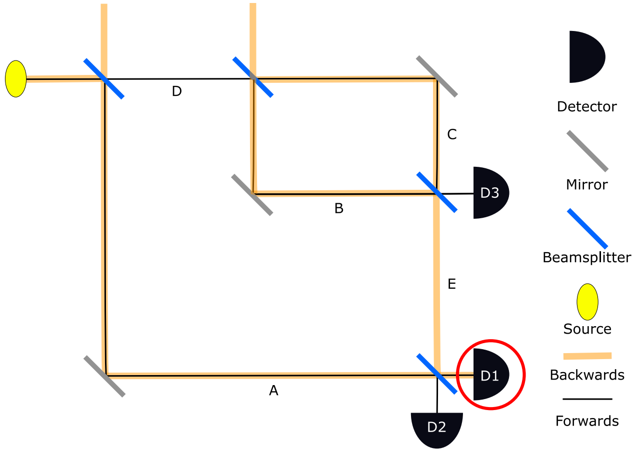

FIG. 1. The Two-State Vector Formalism applied to a nested interferometer (of Salih et al's early type). Forwards-travelling paths are marked by thin black lines, and backwards-travelling paths by thick orange lines. Though no forward or backwards travelling path goes from the source, to Bob (along path C) and into D2, they do overlap over C, meaning there is a weak trace at Bob. This illustrates the peculiar property of the TSVF where particles can jump between regions (e.g. between the inner interferometer and the outer arm) [34].

particle's path, this leads to peculiarities - such as particles jumping discontinuously between locations [73]. This caused Sokolovski to doubt the formalism, in favour of continuous paths. [74–76]. However, these peculiar results don't contradict standard quantum theory [77].

A more compelling counterargument is that the TSVF ignores the non-$\mathcal{O}(\epsilon)$ weak trace, and so does not give the particle's entire path. Vaidman admits this, saying the TSVF gives only $\mathcal{O}(\epsilon)$ elements of the weak trace. Further, analysis of Danan et al's data shows smaller, $\mathcal{O}(\epsilon^2)$ peaks not visible in their original presentation [78].

Vaidman explains this by saying the non-local trace on any particle is also of $\mathcal{O}(\epsilon^2)$, and so this applies in any set-up, even if objects are physically separated. Further, as 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 [34]. Despite this, Vaidman still claims this second-order trace shows a weak trace in Salih et al's most recent protocol [4].

## IV. QUANTUM AS NON-CLASSICAL

Next, we need a criterion for something being quantum, to let us evaluate if the protocols are.

To do this, first consider the differences between classical and quantum physics. For optics (which all protocols so far have used), classical physics consists of everything up to and including Maxwell's equations. These formulate light as the evolution of waves - where possible energies and momenta are continuous [79].

Opposite to this are photons [80], which must be detected discretely, and have energies which are half-integer multiples of a constant, $\hbar$. However, they still have wave-like properties (like interference, when not observed [68]). They is simultaneously wave- and particle-like. Therefore, a protocol is quantum if it only works when dealing with particle-like features. This is typified by single photon detection, where the ability of light intensity to split across different detectors is nullified.

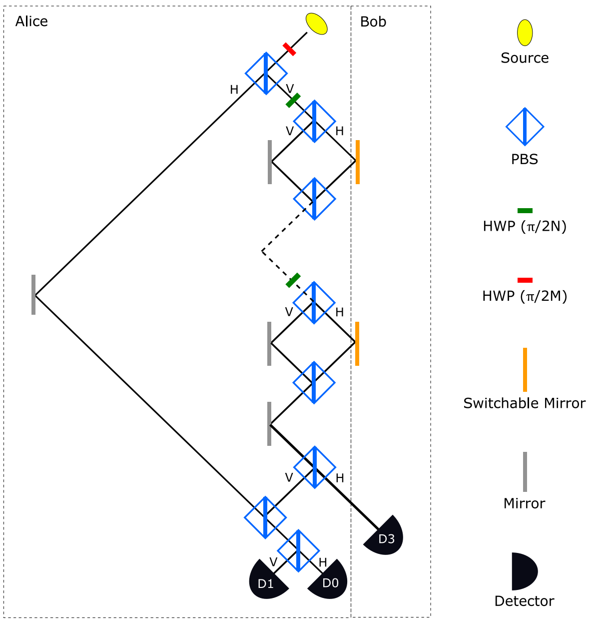

FIG. 2. Salih et al's protocol, shown for one outer cycle. A $H$-polarised photon enters the device, and has its polarisation rotated by $\pi/2M$ (for either $M$ outer cycles, or an $M$ chosen to maximise accuracy if one outer cycle). $H$-polarised elements go through the PBS, and along the outer arm; $V$-polarised elements enter the inner interferometer chain. Again, polarisation is rotated (by $\pi/2N$), and passed through a PBS - $H$-elements are sent to Bob, $V$-elements stay at Alice. If Bob blocks, these elements stay on Alice's side (for infinite inner interferometers, else it has a chance of Bob's blockers absorbing it), so can reach Alice's $D1$; if Bob does not block, elements in the inner chain are sent to a loss channel, so only Alice's $H$ element reaches her detector. The only way Alice's $D1$ can click is if Bob blocks; and in the infinite limit of outer cycles, the only way her $D0$ can click is if he doesn't.

## V. PROTOCOL EVALUATION

Of the protocols proposed so far, only one is counterfactual by both Weak Trace and Consistent Histories approaches - that of Salih et al (Fig 2). Therefore, we want to see if this protocol is quantum. We previously defined quantum as irreplicable by classical physics - meaning its results are unobtainable using classical light (e.g. a coherent state).

In the quantum case, a beamsplitter split the photon's probability of going in either direction; in the classical case, the beam intensity (and field) is split. As interference is still the same, when Bob does not block, waves on both sides still negatively interfere, so the light never returns to Alice. However, Bob's $D_3$ and Alice's $D_0$ both detect light simultaneously. Similarly, when he blocks, light goes to his blockers and Alice's $D_1$ simultaneously. Therefore, in both cases, as light goes between Alice and Bob, it is not counterfactual.

The only way to avoid this is to force the light to end at only one point - to postselect, with information only travelling when nothing goes between Alice and Bob. This can only be done using single photons. Therefore, the only way to make the protocol counterfactual is to use single photons, which makes the protocol quantum.

## VI. CONCLUSION

We have shown Quantum Counterfactual Communication is quantum. This confirms quantum particles are necessary for schemes where we send both bit-values counterfactually, rather than just one of the two. In all schemes proposed so far, this is the only way it is quantum - barring a few tentative papers on the topic, even the information being sent is classical.

However, this 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 [81] - in a new and exciting way, and will hopefully motivate new thought experiments based around this seemingly nonsensical phenomenon.

*Acknowledgements* - We thank Hatim Salih, Will McCutcheon, Paul Skrzypczyk and Robert Griffiths for useful discussions. This work was supported by the UK's Engineering and Physical Sciences Research Council (Grants EP/T001011/1 and EP/L024020/1).

---

[1] R. Penrose, *Shadows of the Mind*, Vol. 4 (Oxford University Press Oxford, 1994).
[2] N. Gisin, Phys. Rev. A **88**, 030301 (2013).
[3] R. B. Griffiths, Phys. Rev. A **94**, 032115 (2016).
[4] L. Vaidman, Physical Review A **99**, 052127 (2019).
[5] T.-G. Noh, Phys. Rev. Lett. **103**, 230501 (2009).
[6] G.-C. Guo and B.-S. Shi, Physics Letters A **256**, 109 (1999).
[7] Y. Sun and Q.-Y. Wen, Physical Review A **82**, 052318 (2010).
[8] Z.-Q. Yin, H.-W. Li, W. Chen, Z.-F. Han, and G.-C. Guo, Physical Review A **82**, 042335 (2010).
[9] M. Jiang, S. Sun, and L. Liang, Journal of Quantum Information Science **1**, 116 (2011).
[10] M. Ren, G. Wu, E. Wu, and H. Zeng, Laser Physics **21**, 755 (2011).
[11] G. Brida, A. Cavanna, I. P. Degiovanni, M. Genovese, and P. Traina, Laser Physics Letters **9**, 247 (2012).
[12] Y. Liu, L. Ju, X.-L. Liang, S.-B. Tang, G.-L. S. Tu, L. Zhou, C.-Z. Peng, K. Chen, T.-Y. Chen, Z.-B. Chen, and J.-W. Pan, Phys. Rev. Lett. **109**, 030501 (2012).
[13] Z.-Q. Yin, H.-W. Li, Y. Yao, C.-M. Zhang, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, Physical Review A **86**, 022313 (2012).
[14] S. Zhang, J. Wang, and C.-J. Tang, EPL (Europhysics Letters) **98**, 30012 (2012).
[15] S. Zhang, J. Wang, and C.-J. Tang, Chinese Physics B **21**, 060303 (2012).
[16] A. Shenoy, R. Srikanth, and T. Srinivas, AQIS 2013 Proceedings (2013).
[17] A. Shenoy-Hejamadi, R. Srikanth, and T. Srinivas, EPL (Europhysics Letters) **103**, 60008 (2013).
[18] Y.-B. Li, Q.-Y. Wen, and Z.-C. Li, arXiv preprint arXiv:1312.1436 (2013).
[19] J.-L. Zhang, F.-Z. Guo, F. Gao, B. Liu, and Q.-Y. Wen, Phys. Rev. A **88**, 022334 (2013).
[20] Y.-B. Li, Quantum information processing **13**, 2325 (2014).
[21] X. Liu, B. Zhang, J. Wang, C. Tang, J. Zhao, and S. Zhang, Physical Review A **90**, 022318 (2014).
[22] A. Shenoy-Hejamadi., R. Srikanth, and T. Srinivas, Phys. Rev. A **89**, 052307 (2014).
[23] A. Shenoy-Hejamadi and R. Srikanth, Phys. Rev. A **92**, 062308 (2015).
[24] T.-Y. Wang, Y.-P. Li, and R.-L. Zhang, International Journal of Theoretical Physics **55**, 5331 (2016).
[25] X. Yang, K. Wei, H. Ma, S. Sun, Y. Du, and L. Wu, Physics Letters A **380**, 1589 (2016).
[26] Y.-Q. Song and L. Yang, arXiv preprint arXiv:1709.08490 (2017).
[27] H. A. Shenoy and R. Srikanth, arXiv preprint arXiv:1311.7127 (2013).
[28] D. R. M. Arvidsson-Shukur and C. H. W. Barnes, Phys. Rev. A **94**, 062303 (2016).
[29] D. R. M. Arvidsson-Shukur, A. N. O. Gottfries, and C. H. W. Barnes, Phys. Rev. A **96**, 062316 (2017).
[30] D. R. M. Arvidsson-Shukur and C. H. W. Barnes, Phys. Rev. A **99**, 060102 (2019).
[31] I. A. Calafell, T. Strömberg, D. Arvidsson-Shukur, L. Rozema, V. Saggio, C. Greganti, N. Harris, M. Prabhu, J. Carolan, M. Hochberg, T. Baehr-Jones, D. Englund, C. Barnes, and P. Walther, npj Quantum Information **5**, 61 (2019).
[32] H. Salih, Z.-H. Li, M. Al-Amri, and M. S. Zubairy, in *The Rochester Conferences on Coherence and Quantum Optics and the Quantum Information and Measurement meeting* (Optical Society of America, 2013) p. T4B.1.
[33] H. Salih, Z.-H. Li, M. Al-Amri, and M. S. Zubairy, Phys. Rev. Lett. **110**, 170502 (2013).
[34] L. Vaidman, Phys. Rev. A **87**, 052104 (2013).
[35] L. Vaidman, Phys. Rev. Lett. **112**, 208901 (2014).
[36] L. Vaidman, International Journal of Quantum Informa-

tion **14**, 1640012 (2016).
[37] H. Salih, Z.-H. Li, M. Al-Amri, and M. S. Zubairy, Phys. Rev. Lett. **112**, 208902 (2014).
[38] R. B. Griffiths, Phys. Rev. A **95**, 066102 (2017).
[39] H. Salih, Phys. Rev. A **97**, 026101 (2018).
[40] R. B. Griffiths, Phys. Rev. A **97**, 026102 (2018).
[41] H. Salih, arXiv preprint arXiv:1807.06586 (2018).
[42] H. Salih, W. McCutcheon, J. Hance, P. Skrzypczyk, and J. Rarity, arXiv preprint arXiv:1806.01257 (2018).
[43] Y. Cao, Y.-H. Li, Z. Cao, J. Yin, Y. Chen, X. Ma, C.-Z. Peng, and J.-W. Pan, in *CLEO: QELS_Fundamental Science* (Optical Society of America, 2014) pp. FM4A–6.
[44] Q. Guo, L.-Y. Cheng, L. Chen, H.-F. Wang, and S. Zhang, Optics express **22**, 8970 (2014).
[45] Q. Guo, L.-Y. Cheng, L. Chen, H.-F. Wang, and S. Zhang, Physical Review A **90**, 042327 (2014).
[46] H. Salih, Physical Review A **90**, 012333 (2014).
[47] S. Zhang, B. Zhang, and X.-t. Liu, arXiv preprint arXiv:1410.2769 (2014).
[48] C. Shukla and A. Pathak, Quantum Information Processing **13**, 2099 (2014).
[49] M. S. Zubairy, Z. Li, M. D. Al-Amri, and H. A. Salih, "Method and apparatus for direct counterfactual quantum communication," (2014), uS Patent 8,891,767.
[50] Y. Chen, X. Gu, D. Jiang, L. Xie, and L. Chen, International Journal of Modern Physics B **29**, 1550134 (2015).
[51] Y. Chen, X. Gu, D. Jiang, L. Xie, and L. Chen, Optics express **23**, 21193 (2015).
[52] Q. Guo, L.-Y. Cheng, L. Chen, H.-F. Wang, and S. Zhang, Scientific reports **5**, 8416 (2015).
[53] M. Al Amri, Z.-H. Li, and M. S. Zubairy, in *Second International Seminar on Photonics, Optics, and Its Applications (ISPhOA 2016)*, Vol. 10150 (International Society for Optics and Photonics, 2016) p. 101501N.
[54] Y. Chen, D. Jiang, X. Gu, L. Xie, and L. Chen, J. Opt. Soc. Am. B **33**, 663 (2016).
[55] H. Salih, Frontiers in Physics **3**, 94 (2016).
[56] Y. Cao, Y.-H. Li, Z. Cao, J. Yin, Y.-A. Chen, H.-L. Yin, T.-Y. Chen, X. Ma, C.-Z. Peng, and J.-W. Pan, Proceedings of the National Academy of Sciences (2017), 10.1073/pnas.1614560114.
[57] Q. Guo, S. Zhai, L.-Y. Cheng, H.-F. Wang, and S. Zhang, Phys. Rev. A **96**, 052335 (2017).
[58] Q. Guo, L.-Y. Cheng, H.-F. Wang, and S. Zhang, Optics express **26**, 27314 (2018).
[59] H. Salih, Royal Society open science **5**, 171250 (2018).
[60] C. Liu, J. Liu, J. Zhang, and S. Zhu, Optics express **26**, 2261 (2018).
[61] F. Zaman, Y. Jeong, and H. Shin, Scientific reports **8**, 14641 (2018).
[62] J. Hance, W. McCutcheon, P. Yard, and J. Rarity, in *Quantum Information and Measurement (QIM) V: Quantum Technologies* (Optical Society of America, 2019) p. T5A.50.
[63] Y. Aharonov and L. Vaidman, Phys. Rev. A **99**, 010103 (2019).
[64] S. Zhang, Chinese Physics B **26**, 020304 (2017).
[65] A. C. Doyle, *The Memoirs of Sherlock Holmes* (George Newnes, 1894) Chap. Adventure 1: "Silver Blaze.
[66] T. Maudlin, *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics* (Blackwell, 2002).
[67] A. Peres, American Journal of Physics **46**, 745 (1978).
[68] T. Young, Philosophical Transactions of the Royal Society of London Series I **94**, 1 (1804).
[69] R. B. Griffiths, Journal of Statistical Physics **36**, 219 (1984).
[70] Y. Aharonov, D. Z. Albert, and L. Vaidman, Phys. Rev. Lett. **60**, 1351 (1988).
[71] B. Tamir and E. Cohen, Quanta **2**, 7 (2013).
[72] A. Danan, D. Farfurnik, S. Bar-Ad, and L. Vaidman, Physical review letters **111**, 240402 (2013).
[73] Y. Aharonov, E. Cohen, A. Landau, and A. C. Elitzur, Scientific reports **7**, 531 (2017).
[74] D. Sokolovski, Quanta **2**, 50 (2013).
[75] B.-G. Englert, K. Horia, J. Dai, Y. L. Len, and H. K. Ng, Phys. Rev. A **96**, 022126 (2017).
[76] D. Sokolovski and E. Akhmatskaya, Annals of Physics **388**, 382 (2018).
[77] U. Peleg and L. Vaidman, arXiv preprint arXiv:1805.12171 (2018).
[78] D. Sokolovski, Physics Letters A **381**, 227 (2017).
[79] D. J. Griffiths, "Introduction to electrodynamics," (2005).
[80] A. Einstein, Annalen der Physik **322**, 132 (1905).
[81] J. R. Hance, arXiv preprint arXiv:1909.06608 (2019).
[82] A. C. Elitzur and L. Vaidman, Foundations of Physics **23**, 987 (1993).
[83] P. Kwiat, H. Weinfurter, T. Herzog, A. Zeilinger, and M. A. Kasevich, Phys. Rev. Lett. **74**, 4763 (1995).

## Appendix: Quantum Counterfactual Communication Proposals

In this Appendix, we look at Quantum Counterfactual Communication protocols proposed so far. Since Elitzur and Vaidman first discovered quantum counterfactuality [82], and Kwiat et al allowed loss to be made effectively nil [83], researchers have tried to exploit it for communication. Despite this, all protocols until recently have fallen into three broad categories: where communication is counterfactual only for one bit-value; where photons travel between Alice and Bob, but in the opposite direction to the information passed between them; and where no photons pass between Alice and Bob when information flows, but the error/loss rates vary with the bit-value Bob sends.

### 1. Elitzur-Vaidman Bomb Tester

All quantum counterfactual communication protocols stem from the Elitzur-Vaidman Bomb-Detector [82]. In this thought-experiment (Fig. 3), a balanced MZI has a potentially faulty bomb along one of its paths, which can only be detonated by a non-demolition single-photon detection.

If the photon goes along the bomb's side of the MZI (and the bomb works), it detonates, and the photon (and everything else) is destroyed; if the bomb is faulty, the photon travels to the merging beamsplitter normally. However, if the photon travels along the other side, the bomb working changes the interference pattern, making

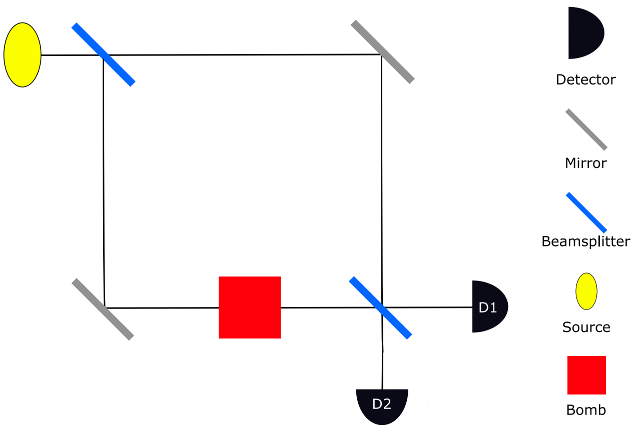

FIG. 3. The Elitzur-Vaidman Bomb Tester. A photon is emitted from the source (top-left), enters the balanced Mach-Zehnder interferometer, and is spread across both paths equally. If the bomb is faulty, the photon recombines at the second beam-splitter, and enters detector 1 with 100% probability. If the bomb works, and is activated, the entire apparatus is destroyed. If the bomb would work, but the photon went down the bomb-free path, the photon has a 50:50 chance of being detected at each detector.

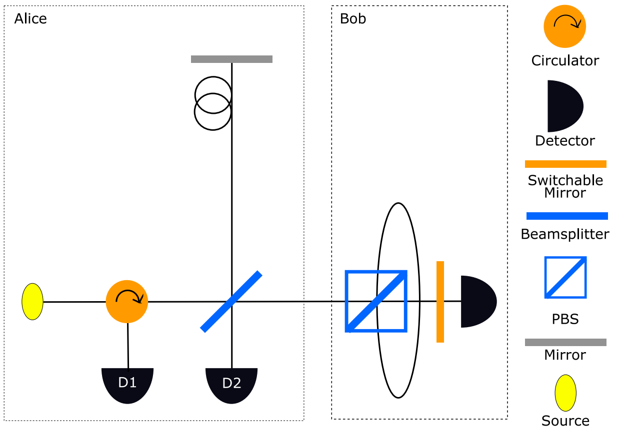

FIG. 4. Noh's counterfactual cryptography protocol - Alice randomly polarises a photon, which passes through a beam-splitter, with one of the outputs going to Bob. There, if the photon is orthogonal to Bob's choice of polarisation, it passes through, is reflected back, and interferes in Alice's interferometer, going to $D_2$. However, if the polarisation matches Bob's choice, the wave is sent into his detector. If this clicks, the protocol is aborted; if not, this removes the interference, and forces the photon on Alice's side into her $D_1$ [5].

it able to go to a detector it previously couldn't access. This 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. Unlike classical counterfactual communication, both options (the bomb working and the bomb not working) are transmitted counterfactually - a valid path can be drawn from the source to the detectors without going via the object (bomb) under evaluation. However, it is also unlike it as it is not necessarily always counterfactual. This is as, while the photon can carry the information without going via the bomb, it does not necessarily have to. This means the Bomb-Tester is not fully counterfactual.

### 2. Counterfactual only for One Bit

In the next form of Quantum Counterfactual Communication, the protocol is counterfactual for one bit-value, but the photon goes between Alice and Bob for the other.

The first of this type of protocol, and indeed the first formal Quantum Counterfactual Communication protocol proposed was Noh's (Fig. 4), published in 2009 [5] (barring Guo's adaption of the E-V Bomb Detector, where photons travel between Alice and Bob for both bit-values [6]). For matched polarisations, if Alice gets a click, the photon has remained on her side. However, for orthogonal polarisations, it has both been to, and returned from, Bob - so is not counterfactual.

Despite this, the work generated a lot of interest, with security protocols and analyses based on it still being generated [7–27]. While plenty of these focus on reducing loss by reducing the proportion of the photon sent to Bob [7], this must always be non-zero for the protocol to function. Therefore, the system will always have some non-counterfactuality.

### 3. Information and Photon Travel in Opposite Directions

The next category is where the photon can cross the channel, but does so in the opposite direction to the information being sent. Here, for one bit-position the photon destructively interferes across the quantum channel, keeping it at Alice, and for the other, it constructively interferes, allowing it through to Bob. Based on if she detects a photon, Alice can determine which bit Bob sent.

The only protocol of this sort is Arvidsson-Shukur et al's. They propose a device formed of chained MZIs, which use the Quantum Zeno effect to (for many MZIs) keep the photon at Alice if Bob blocks, and force it to go to Bob if he does not [28]. This is identical to Kwiat et al's Interaction Free Measurement protocol [83].

Ignoring the high chance of Alice wrongly believing Bob did not block (due to the necessarily finite number of MZIs causing some chance of a blocker absorbing the photon), there is still the issue that the photon travels at the same time as the information. Waves carrying information in the opposite direction to travel is a well-known classical phenomenon [4]. Therefore, this protocol only seems quantum when you consider light as local - where, for one possibility, the photon travels from Alice to Bob. This obviously creates a weak trace at Bob, and so is not counterfactual.

Arvidsson-Shukur et al attempt to advocate their protocol by saying it tolerates error better than others [29], and by calling other protocols classical using a classical model with Alice and Bob having extra, non-trivial resources (e.g. a shared clock) [30]. However, they give no reason to view their protocol as true Quantum Counterfactual Communication, and so their work with Calafell et al [31] just demonstrates classical counterfactuality.

### 4. Photon only Travels Erroneously - Unequal Losses

The next set of protocols is where Alice receives a photon for both bit values, which has never been to Bob. This means the photon cannot go to Bob when Alice gains information, as then Alice would be unable to see which bit was sent. Therefore, when photons go to Bob, the protocol must be aborted and retried, creating a source of loss in the system.

For these protocols, this loss varies with the bit-value sent. This leads us to ask if Alice can, by knowing loss probabilities, guess the bit Bob sends solely based on if any of her detectors light up. This would make the same as the last category. We can also ask, even with this loss enforced according to the protocol, if this post-selection is to blame for any peculiar effects observed.

Salih et al's 2013 protocol was the first claiming to be fully counterfactual for both bit values [32, 33]. It is formed of a chain of outer interferometers, each containing a chain of inner interferometers (see Fig.2).

However, Vaidman claimed this was not counterfactual, when assessed by the Weak Trace criterion (see Fig. 1) [34]. The TSVF gives a weak trace on Bob's side of the channel when Bob does not block, meaning we cannot say the photon was not there [35, 36]. However, this is only for the simplest (polarisation-free) form of the protocol. Using polarisation (as shown in Fig.2 and [42]) avoids a weak trace on Bob's side by ensuring the only waves that go to Bob are H-polarised, which are lost via $D_3$ on Bob's side, restarting the protocol [37]. Then, when Bob does not block, the differences in polarisation between the forward- and backward-travelling states keep them separate on Bob's side - giving no Weak Trace there. This was tested practically through weak measurement, using Danan et al's method [72] and shown to have no weak trace from Bob's side visible at Alice's detectors.

Griffiths also claimed Consistent Histories shows it as un-counterfactual, as a history with a non-zero probability could be traced to Bob's side and back when Bob does not block [3, 38, 69]. However, again, Griffiths only considered physical paths, rather than polarisations, which provide an extra degree of modal separation [39]. Griffith later claimed, when using more than one outer cycle, the family became inconsistent, and so it was meaningless to say the protocol was counterfactual [40]. However, as Salih notes, the final cycle of a chain is counterfactual, while the identical earlier ones are meaningless, which seems paradoxical [41].

Once Salih et al published their protocol, various implementations began to appear [23, 43–61]. While many of these don't make use of polarisation, some do, alongside wider modal-style analyses of the protocol [41, 62].

Despite originally claiming counterfactual communication of both bit-values was impossible, at roughly the same time as Salih et al defended their protocol using polarisation, Vaidman, alongside Aharonov, released a protocol allowing just this [63]. This method is effectively the same as in Salih et al's original protocol - however, to avoid a weak trace in this set-up, where there is no 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 the protocol; but all this does is connect the two inner interferometers, and fold the outer path to reduce physical space used, and so it is irrelevant to the protocol's counterfactuality. However, unlike Salih et al's, it is not counterfactual by the Consistent Histories approach.

Zhang et al proposed a protocol, based on Salih et al's original method, for probabilistic counterfactual communication. They admit their protocol isn't always counterfactual, but claim the chance of the photon being at Bob can be reduced to practically nil, and losses (from noise and blocking) are reduced [64]. However, they base their claim that this protocol is counterfactual most of the time on the assumption the photon only traces one path, rather than Consistent Histories or Weak Values - it is not counterfactual.

### 5. Photon only Travels Erroneously - Equal Errors

Shortly after publishing with Aharonov, Vaidman created another weak trace-free counterfactual communication protocol. However, for one outer cycle, this protocol avoids the risk of an erroneous reading that Salih et al's, and his earlier, protocol has [4]. It is again based on a chained MZI set-up, but uses 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 it is what Bob sent. Like his protocol with Aharonov, it requires more than one inner interferometer to be fully counterfactual, as otherwise the lack of the polarisation degree of freedom means a weak trace appears on Bob's side by the TSVF. However, another benefit of the protocol is that, for certain beam-splitting values in the two inner interferometer case, losses were the same whether or not Bob blocked. This means Alice cannot infer if Bob blocked, just based on if she receives a photon. Therefore, the protocol cannot be reduced to the information and photon travelling simultaneously, in opposite directions, and so, by the Weak Trace approach, it is counterfactual. However, again, by the Consistent Histories approach, it is not counterfactual.