QUANTUM COUNTERFACTUALS: *INTERACTION-FREE MEASUREMENTS* AND CLASSICAL LOGIC'S CONSTRAINTS

Carmen SÁNCHEZ-OVCHAROV, Karim GHERAB-MARTÍN, J.A.F. CUESTA

ABSTRACT: The aim of this paper is to carry out a critical analysis of the uses of the material conditionals employed to represent the counterfactual detections taking place inside Elitzur-Vaidman quantum bomb detectors in so-called quantum interaction free measurements (IFMs). We provide an exhaustive philosophical study of this classical approach focusing on its limitations. To do so, we formulate atomic propositions corresponding to the observable events of the experiments, to overcome the shortcomings of the conditional expression. Finally, we show that the standard definition of counterfactuality in quantum contexts based on classical logic is epistemologically too limited, presenting quantum logics as the most promising possible approach to address this type of quantum counterfactuality.

KEYWORDS: Interaction-Free measurements, counterfactuals, quantum mechanics, conditionals

1. Introduction

Quantum counterfactuality is an interdisciplinary field that brings together quantum physics, mathematics, the philosophy of physics, and logic. More recently, quantum computing has also entered the picture, given the potential computational applications of counterfactuality.[1] The theoretical core of this interdisciplinary field originates from a thought experiment proposed by Avshalom Elitzur and Lev Vaidman in 1993: the so-called "quantum bomb detector problem". The thought experiment raises the possibility of performing detection measurements without interaction with the object to be detected ("Interaction Free Measurements", or IFMs). Their interest is not only relevant within the quantum computational community, IFMs also provide the ideal context to revisit the most intense debates surrounding the interpretation of quantum mechanics, and specifically the opposing Everett's many-worlds and orthodox or Copenhagen interpretation. This is because

[1] Elitzur A. & Vaidman L. (1993) have already raised the possibility of *counterfactual quantum computing.* In the following investigations stand out: Mitchison & Jozsa (2001), Mèthot & Wicker (2001), Azuma (2004), Hosten et al. (2006), Vaidman (2007), and Raj et al. (2019).

those who have proposed and developed the theoretical core, especially Vaidman (2018), have used the counterfactuals found in the IFM frameworks, as epistemological and experimental justification for the *many-worlds* interpretation.

A rigorous evaluation of the issues requires a logical and philosophical clarification of the *formalization* of the process of counterfactual detection which all these issues revolve around (Cuesta and Sánchez-Ovcharov 2023, 333). Our aim in this paper is precisely to carry out such an exercise in logical and philosophical clarification by entering some of the main results in so-called quantum logics and classical semantics.

The first part of this paper is devoted to an exposition of the operation of the two devices that allow us to speak of "counterfactual detections" and "measurements without interaction": the Mach-Zehnder interferometer and the Elitzur-Vaidman bomb detector. As we shall show, the latter is based upon the former. The second part of the paper introduces the traditional formalization, in terms of classical logic, which has been given to the counterfactual detection process. Through a simple formal expression of two propositions articulated by a material conditional, we show its insufficiency. In the third and last part of the paper, we explore other approaches using classical connectives and their limits to build a formal articulation of all the processes involved in counterfactual detections. We expose the best classical alternatives and analyze their epistemological limits. Following on these results, we raise several questions on the suitability of this classical approach present in the current academic literature and point out some possible ways to connect quantum counterfactuals and study them from a logical perspective. We finally mention, in a nutshell, standard quantum logics based on Hilbert space lattices as the most promising epistemological approach to quantum counterfactuality.

2. The Mach-Zehnder interferometer: Definition and function

Mach-Zehnder interferometry devices (Grynberg et al. 2010) were originally developed to analyze phase shifts between suitably collimated light beams coming from the same light signal emitted at the source. This setup consists of a light signal source, two mirrors (M1 and M2), two detectors (D1 and D2) and two beam splitters (BS1 and BS2). In a quantum context this means ½ probability for the photon to be reflected and ½ to be transmitted. The complete device is shown in the figure below:

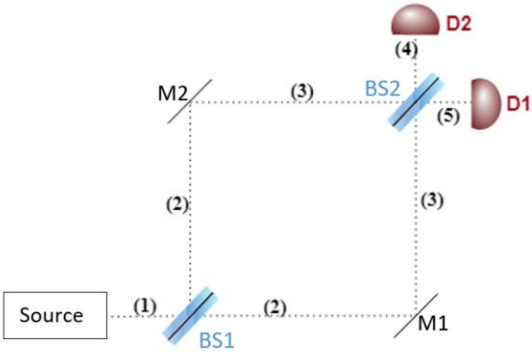

Figure 1. Schematic diagram of Mach-Zehnder interferometer: light signal (Source), two mirrors (M1 and M2), two detectors (D1 and D2) and two beam splitters (BS1 and BS2).

When the phenomenon of reflection of the light beam takes place, either on the half-silvered surface of a beam splitter (BS1, BS2) or directly onto one of the mirrors (E1, E2), a phase shift is introduced with respect to the initial beam's phase.[2] If we consider an incident monochromatic wave, a reflection at a beam splitter introduces a phase shift of $\pi/2$ in the reflected component. This phase shift must not be understood as a spatial displacement of $\lambda/4$, but rather as a relative phase difference between the transmitted and reflected amplitudes. It is precisely this phase relation —and not a physical path difference— that determines the constructive and destructive interference patterns observed at the output ports of the interferometer.

Suppose that the light source is a precision device that can send photons each at a time.[3] Hence a single photon is traveling along one of the device's paths at any time. Yet, the resulting phenomenon is not as simple as it would be from a classical-corpuscular view of the experiment. As in the famous double-slit experiment, the photon seems heuristically to somehow interfere with itself, as follows.

We represent as $|\rightarrow\rangle$ in the Dirac's notation the state of the photon traveling to the right or *right state* and as $|\uparrow\rangle$ the state of the photon traveling upward or *upward state* (inside the structure represented in Figure 1). This also allows us to introduce the standard quantum-logical algebraic approach. The origin of standard quantum logics due to Birkhoff and von Neumann (1936) has undergone many

[2] The value of the phase difference is not a big deal for the explanation at hand, besides, the phase difference in mirrors M1 and M2, being equal, will not affect the phase difference, and could be included in a global phase difference.

[3] Many of the discussions regarding whether these experiments are possible under optimal conditions depend upon the possibility of emitting photons one at a time.

developments we cannot explore here. However, most of them have emerged from approaches based on different non-classical semantics[4] aimed at precisely capturing certain results or striking aspects of quantum mechanics. In the end, we will say something about it. Strictly speaking, standard quantum logics are in a nutshell nothing more than the logical interpretation of the algebra that can be generated from the class of linearly closed subspaces over Hilbert space (*H*). This structure is a non-distributive (i.e. non-Boolean) orthomodular lattice isomorphic to the set P(*H*) of all the projector operators defined over the Hilbert space. And the quantum-states' mathematical representations (each closed subspace) are taken as *quantum* propositions (Svozil 1998). Moreover, the representation of quantum systems through these subspaces is precisely what we express using the bra-ket notation to directly denote the inner product operation on *H* following the standard approach.

The photon's successive reflections can then be represented as follows:[5]

$$|\rightarrow\rangle \Rightarrow_{(Reflection)} e^{i\pi/2}|\uparrow\rangle = i|\uparrow\rangle \Rightarrow_{(Reflection)} -|\rightarrow\rangle \quad (1)$$

This expression should be read as a state evolution, whereby the initial right state of the photon $|\rightarrow\rangle$, undergoes a reflection, indicated by a phase shift $e^{i\pi/2}|\uparrow\rangle$, and thus becomes the upward state $i|\uparrow\rangle$, but after a second reflection returns to a right stat*e* $|\rightarrow\rangle$.

$$|\uparrow\rangle \Rightarrow_{(Reflection)} e^{i\pi/2}|\rightarrow\rangle = i|\rightarrow\rangle \Rightarrow_{(Reflection)} -|\uparrow\rangle \quad (2)$$

This expression should be read as another physical evolution that takes the initial upward state of the photon$|\uparrow\rangle$ which, after reflection and phase shift, ends up in the upward state $i|\uparrow\rangle$.

However, in the physical transmission event there is no obstacle in the photon's path, and consequently no real phase shift. The photon rather seems to interfere somehow with itself, so that the interference detected at D1 is a constructive one, while that detected at D2 is destructive. This is a dramatic consequence of forcing classical terms as *trajectory* in quantum devices. To arrive at these conclusions, let us record the photon states in each section of the interferometer as a superposition of states. Here each state can be identified with a quantum proposition:

---

[4] For a general companion of different philosophical and technical discussions including different non-classical semantical approaches we remit to Dalla Chiara and Giuntini (2002); Weingartner (2004) and Engesser, et al. (2007).

[5] Notice that all the reflections that take place in the Mach-Zehnder interferometer produce a phase shift of π/2. Note also that the minus sign of the state argument in the second reflection of a photon coming from the light source represents a global phase that will be indifferent to us since it has no physical meaning. Quantum observables rely on the squared modulus of the amplitudes.

Photon leaving the light source:	$\|\rightarrow\rangle$	(3)
Photon after impact on BS1:	$\frac{1}{\sqrt{2}}(i\,\|\uparrow\rangle + \|\rightarrow\rangle)$	(4)
Photon after reflection in M1 and M2:	$\frac{1}{\sqrt{2}}(-\|\rightarrow\rangle + i\|\uparrow\rangle)$	(5)
Photon after impact in BS2:	$-\|\rightarrow\rangle$	(6)
$\|\rightarrow\rangle$ state after BS2:	$\frac{1}{\sqrt{2}}(\|\rightarrow\rangle + i\|\uparrow\rangle)$	(7)
$\|\uparrow\rangle$ state after BS2:	$\frac{1}{\sqrt{2}}(\|\rightarrow\rangle + i\|\uparrow\rangle)$	(8)

These characteristics of the Mach-Zehnder interferometry device are the basis of the Elitzur-Vaidman quantum bomb detector, which is why Vaidman (2008) writes that "the simplest EV device is the Mach-Zehnder interferometer". Let us now look at the characteristics of the Elitzur-Vaidman device: the so-called *quantum bomb detector*. This will introduce the quantum counterfactuals.

3. The Elitzur-Vaidman bomb detector: interaction free measurements (IFMs) and counterfactuals

The so-called *Elitzur-Vaidman quantum bomb tester* (Elitzur & Vaidman, 1993) only adds one structural element to the Mach-Zehnder interferometer, an opaque object, e.g., a *bomb*, which goes off when a photon, incident at BS1 and traveling down the M1 path of the interferometer, strikes it. This minimal modification at the assembly level causes the interference pattern to be lost, again in analogy to the famous double-slit experiment.

Diagrammatically, the EV device can be represented as follows:

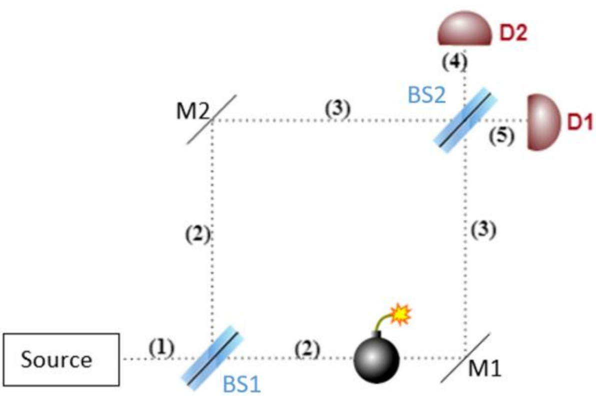

Figure 2. Schematic representation of the Elitzur-Vaidman bomb tester: light signal source, mirrors M1 and M2, detectors D1 and D2, beam splitters BS1 and BS2, and an opaque object, the bomb.

The recording of the successive photon states are the same as in the Mach-Zehnder interferometer, except for a new state for the photon, called $|boom\rangle$. This represents the state when it hits the bomb located between BS1 and M1 and the bomb explodes. We reformulate the paths and propositions for the EV detector as follows:

Before BS1:	$|\rightarrow\rangle$	(9)
After BS1:	$\frac{1}{\sqrt{2}}\,(|\rightarrow\rangle + \,i\,|\rightarrow\rangle)$	(10)
After M1 and M2:	$\frac{1}{\sqrt{2}}\,(|boom\rangle - \,|\rightarrow\rangle)$	(11)
After BS2:	$\frac{1}{\sqrt{2}}|boom\rangle - \frac{1}{2}|\rightarrow\rangle - \frac{i}{2}|\uparrow\rangle$	(12)

If we pay attention to the photon's states after passing through BS2, in terms of (12), we can determine the three probabilities corresponding to the three possible outcomes as follows:

Probability of achieving $|boom\rangle$:	$\left|\frac{1}{\sqrt{2}}\right|^2 = \frac{1}{2}$	(13)
Probability of achieving $|\rightarrow\rangle$:	$\left|-\frac{1}{2}\right|^2 = \frac{1}{4}$	(14)
Probability of achieving $|\uparrow\rangle$:	$\left|\frac{i}{2}\right|^2 = \frac{1}{4}$	(15)

From Born's rule we obtain the standard Elitzur-Vaidman probabilities for an ideal 50/50 interferometer: ½ for explosion, ¼ for detection at D1, and ¼ for detection at D2. Crucially, the ¼ probability at D1 is not an interferential effect: the presence of the bomb destroys the phase coherence between the arms, so the amplitudes reaching BS2 do not interfere. The port D2 therefore becomes the only genuinely interferential signature of the device, and its non-zero probability is what allows interaction-free detection.

Note, therefore, that the last probability for a final upward state of the photon has changed, relative to the simpler MZ set up, where the bomb is absent. Thus, we can conclude that the presence of the opaque object (bomb) entails a ¼ probability of detecting its presence, without any apparent interaction with it —since there has been no explosion at all.[6]

---

[6] A distinction is usually made between the detection of the existence of an opaque object sensitive to contact with a photon and a bomb whose action device is triggered by a photon detector. Here we choose to treat the *opaque object* as functionally indistinguishable from a photon contact-sensitive bomb. Eliztur and Vaidman claim that "in one respect the experiment which tests a bomb without exploding it is easier than the experiment of testing the existence of an object in a given place without touching it" since "it is possible to obtain certain information about a region in space

Note that we may also remove detector D1, thus leaving the experiment with only two structural elements —bomb and D2—, but with no consequence for our chosen formal representation of the device. The reason is that the photon traveling along the path to D1 is recoverable for the photon source since it is sufficient to close the circuit between D1 and the photon source, as follows:

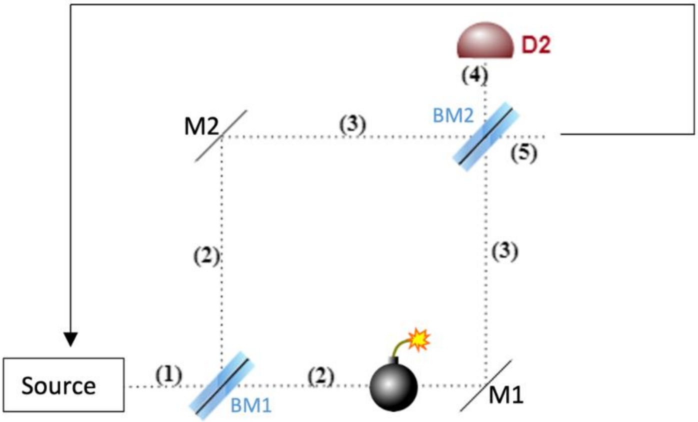

Figure 3. Schematic diagram of Elitzur-Vaidman bomb tester with one detector (D2).

The case in which the photon does not appear to interact with the bomb,[7] there is no explosion, and the photon is subsequently detected at D2, is known as *interaction free measurement* (IFM).[8] It is so named because it allows us to determine the presence of the bomb in the device, without exploding it —i.e. a *counterfactual* detection.

4. Traditional formalization of the counterfactual detection process

The objective of the Elitzur-Vaidman device is to determine the presence of a bomb in a counterfactual manner —that is, by detecting its existence without triggering

without any interaction in that region neither in the past nor at present".

[7] The conditions and terms under which this *interaction* occurs give rise to discussions on interpretations of quantum mechanics, which are beyond the scope of the present work. We use the terms "appear" and "apparent" to remain neutral with respect to such discussions for the time being. Omitting it may give rise to a tacit acceptance of Everett's *Many-worlds* interpretation, to which Vaidman himself subscribes. There is no local interaction of any kind in the *actual world*. The interaction takes place in that possible world in which the bomb would, *in fact*, explode.

[8] See Elitzur and Vaidman (1993), Vaidman (1994, 2003, 2008, 2018, 2019), Penrose (1994), Kwiat *et al.* (1995) and Mitchison & Jozsa (2001).

an explosion. The answer is obtained by detecting the presence of the bomb without exploding it. And the threads that are involved in this detection are collected in the following *flowchart*:

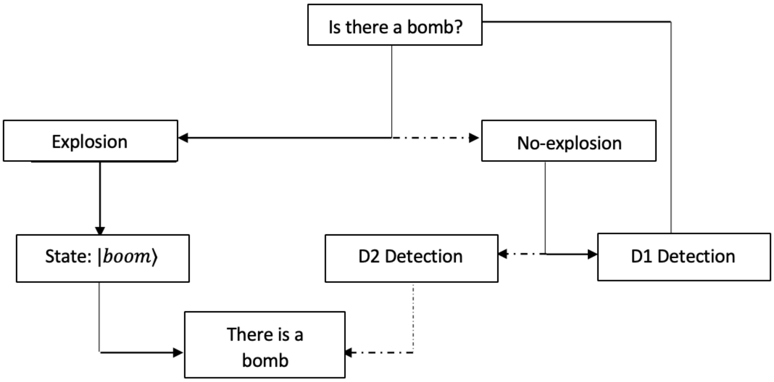

Figure 4: Counterfactual detection procedures represented in a flow diagram. A dotted line shows the 'counterfactual' detection procedure, while the solid lines show 'actual' ones.

As we shall see, Figure 4 represents not only the three possible results (bomb's explosion, D2's detection and D1's detection), but also their *logical* relationship with respect to the overall functioning of the detection procedure.

5. Logical approaches to the counterfactual detection procedures

In philosophy of logic, "counterfactual" usually refers, in the first instance, to the conditional proposed by Stalnaker (1968), whose semantic proposal was later developed by Lewis (1973). Suffice it to note, for the time being, that the notion of "counterfactual" proposed by these two authors is framed within a modal approach that is usually related with mainstream modal semantics. Their purpose does not differ, at least in spirit, from the motivation behind using strict conditionals to model every-day conditional inferences, relevant implications, Cooper's paradoxes, and similar phenomena. Obviously, none of these counterfactuals were developed in response to the need to formalize IFMs.

They have nothing to do with the characterization of counterfactuals —not as mere speculative or conditional informal inferences, but as physical phenomena. It is true that some indirect connections have been drawn. While Lewis himself explicitly linked his semantics for counterfactuals to a multiplicity of real possible worlds[9], and recent works have related this formalism to many-worlds

[9] Lewis (1993) proposed a realism which, on the other hand, is very hard to uphold after

interpretations, we do not consider this kind of modal realism either necessary or appropriate for an initial philosophical approach to IFMs.

Secondly, Vaidman and Elitzur, among others, have also framed their results within the broader set of possible arguments supporting realist quantum interpretations. However, our primary interest here is to characterize IFMs as neutrally as possible. Once we adopt the standard quantum-logical approach and Hilbert space notation, this task becomes particularly demanding.

It is, moreover, quite common in the philosophical literature to understand counterfactuals in the strongest "classical" sense of the term —that is, as the subjunctive form of material implication, as follows. Consider the interpretation of natural language that requires a subjunctive mode: "If it were the case that 'p', then it would be the case that 'q'," where it is implicitly asserted that, in fact, neither 'p' nor 'q' holds. Formally, this is typically represented as: $(p \rightarrow q) \wedge (\neg p \wedge \neg q)$. This conception of counterfactuals as subjunctive or hypothetical traces back to Aristotelian and Megaric-Stoic dialectics (Cuesta 2021), becomes widespread in scholastic discussions, and gains prominence among modern philosophers — especially after Wolff and Kant.

And this the one that has been applied to IFMs. This is the conception that has been taken to hold between the interaction of the photon with the bomb and its explosion, even when there is, in fact, neither impact nor explosion.

Cuesta and Sánchez-Ovcharov (2023) point out that this approximation suffers from a deficiency intrinsic to the formalism itself: a statement of the type $(p \rightarrow q) \wedge (\neg p \wedge \neg q)$ simply expresses a formal logical relation, in which it is shown that a conditional (any) will be true even if antecedent and consequent are false. To overcome this limitation, they introduce a distinction between *informative* and *non-informative* counterfactual results, which demand caution when qualifying a result as properly counterfactual. The isolated hypothetical conditional will in no case be sufficient to distinguish between *informative* and *non-informative* counterfactual statements. And both Mitchison and Jozsa (2001), as well as Cuesta and Sánchez-Ovcharov (2023) when referring to *informative* counterfactual results, understand that counterfactual detection is nothing else than "a detection which results from a procedure represented by a true counterfactual conditional".

incorporating the interpreted Elitzur-Vaidman results.

6. Analysis of the conventional formalization: the subjunctive or hypothetical conditional

As we have already mentioned, the most widespread formalization of the counterfactual detection process has been carried out in classical logic by means of the classical conditional: $p \rightarrow q$. Here, propositions *p* and *q* simply stand for "photon interacts with the bomb" and "the bomb explodes", respectively. In the counterfactual detection process it is commonly taken that the truth values for both, p and q, are 0 or *false*. This means that p as well as q are false propositions while the conditional remains true (Kwiat et al. 1995; Vaidman 2003; Cuesta and Sánchez-Ovcharov 2023) as in the classical truth-table in table 1.

	p	q	$p \rightarrow q$
i1	1	1	1
i2	1	0	0
i3	0	1	1
i4	0	0	1

Table 1: Corresponding to the conditional. The interpretation of the conditional called *counterfactual*, *hypothetical* or *subjunctive* is highlighted.

Vaidman (2003, 495) introduces this conditional by pointing out that "simple logic tells us: given that any interaction leads to an explosion and given that there has been no explosion, it follows that there has been no interaction". This reasoning can be formalized with a *modus tollens* in "simple [classical] logic" as follows — where *p* is "the photon interacts with the bomb" and *q* "the bomb explodes":

$$p \rightarrow q, \neg q \vdash \neg p \qquad (16)$$

Thus, the conditional is included as a premise, i.e., taking it as true — specifically in that case in which it is true, but the antecedent and consequent are both false (*i4*). The logical validity of this argument implicitly requires an equivalence principle,[10] namely the rule of the contraposition of the conditional, formally defined as follows:

$$p \rightarrow q \equiv \neg q \rightarrow \neg p \qquad (17)$$

Where we have used the metalogical symbol, ≡, to represent the relation of logical equivalence. According to this logical rule, the truth of a conditional implies the truth of its contrapositive and *vice versa*.

[10] This equivalence principle can be alternatively regarded as a logical law or a rule of inference; the difference is not relevant for our purposes.

Bearing this in mind, we are already able to see that *i4* is insufficient to represent counterfactual detection processes. The problem is not that the conditional collapses to i4, but that i4 alone does not track any of the physical distinctions relevant to the experiment. The classical conditional simply assigns the value 1 whenever both antecedent and consequent are false, but this semantic fact does not correspond to any physical process occurring inside the interferometer. Therefore, the material conditional cannot distinguish between genuine interaction-free detection and mere logical vacuity.

The two conditionals in (17) are only equivalent by virtue of formalizing the same proposition and, ultimately, the same conditional events. If we assume that the expression "$p \longrightarrow q$"stands for the event that "if the photon impacts, then the bomb explodes", then "$\neg q \longrightarrow \neg p$", by definition, will necessarily only formalize the same fact. But, even if we disregarded the equivalence between the conditional and its contrapositive —which Vaidman fails to do—, *i4* is still not sufficient to formalize counterfactual detection processes.

First, to avoid confusion, we reject the term "counterfactual conditional" for the interpretation of the conditional in which both, the antecedent and the consequent, are false. We instead call it, on the grounds of what has just been explained, simply "*i4*" or just "classical logical counterfactual". We conjecture that the use of the name "counterfactual" for this interpretation of the conditional may have led to the mistaken identification of two different types of "counterfactuality". One of them refers to the formal interpretation of the logical definition of the conditional: the classical one. The other refers to the empirical detection of an object —the bomb— without any interaction with it: the quantum one. It seems that this misidentification may have led to the failure of the formalization of the process of counterfactual detection by means of a single conditional.

Secondly, *i4* does not allow us to study the mathematical and physical events that make the conditional true. On the one hand, calling this interpretation "counterfactual" would have no more implications than calling it "*i4*" or "Alice". It would be a mere question of nomenclature. On the other hand, justifying it on the grounds that the counterfactual relies on the assumption that neither the antecedent nor the consequent are in fact given —they are false— but the conditional is nonetheless true, seems to be an illegitimate extra-logical leap, based on an extrapolation from a logical form to an empirical result.

Let us use an example to illustrate the impossibility of formalizing IFMs by means of *i4* only. We simplify the original thought experiment as follows: we have a black box in which there might be a bomb, and we want to determine if there is one. We can only do so by opening the box and looking inside.[11] We can express the situation in conditional terms as follows: "if I open the box, then I will detect if there is a bomb inside". Whereby we understand "I will detect" in both senses: either "I will detect that there is a bomb" or "I will detect that there is no bomb". Formally, "$p \longrightarrow q$"with p as "I open the box", and q as "I will notice that there is a bomb".

Now take *i4*. We have a conditional that is true when, in fact, we have not opened the box $v(p) = 0$, and in fact, we have not detected a bomb, $v(q) = 0$. To claim that, due to the conditional being true under *i4*, we have truthfully determined or ascertained something counterfactually concerning the bomb or the box is simply to commit a category mistake. We have determined nothing either factually, nor counterfactually. Only the formal truth of the conditional $p \longrightarrow q$ has been maintained. This is what we earlier referred to as an extra-logical leap: an extrapolation from the truth of a logical form to the truth of an empirical event. Delving a little further into it, the conditional itself, in any of its interpretations, neither affirms nor confirms any fact about the material content of its terms. It only establishes the formal truth about a formal expression of a hypothetical nature. In this sense, in the case of *i1*, when we have opened the box and have detected the presence of the bomb —because it has exploded— or we have just detected that the box is empty, in this case also the conditional has a truth value 1 —as in *i4*. But, in addition we have factually determined that there is/is not a bomb in the box, and we have transformed the hypothetical matter into a factual issue. Now it seems clear that the presumed counterfactuality, derived from the extra-logical leap, has nothing to do with that obtained in IFMs because in IFMs a counterfactual detection is, *in fact*, obtained.

What the conditional "$p \rightarrow q$" seems to offer is rather, at best, a *functional definition* of the Elitzur-Vaidman bomb itself, so that each of its interpretations corresponds to a formal description of the different material possibilities of the experiment (Table 2):

	p	q	$\boldsymbol{p \rightarrow q}$	Material possibilities of the experiment
i1	1	1	1	Photon impacts | Bomb explodes
i2	1	0	0	Photon impacts | Bomb does not explode
i3	0	1	1	Photon does not impact | Bomb explodes
i4	0	0	1	Photon does not impact | Bomb does not explode

Table 2. Material possibilities of the Elitzur-Vaidman bomb detector.

[11] For the analogy to be complete, we can add that, if we open the box, and there is a bomb, it will explode, but this is an unnecessary complication in this setting.

We may consider whether this conditional is a correct formalization of the bomb's design: a device such that if a photon, p, hits it, then the bomb explodes, q. But it is necessary to realize that even this definition is incomplete.

The logical form of the conditional includes an interpretation, *i3*, $p \rightarrow q$: $v(p) = 0$ y $v(q) = 1$, which we will reject since it is not applicable to our experiment as described with all its parameters under control. This interpretation would ultimately model at the very most a faulty bomb or failure due to e.g. external causes to the experiment, making it no longer ideal —or not an interpretation at all. By removing *i3* we rule out events such as: the deterioration of the pump, that it interacts with a photon not coming from our light source (even one outside our laboratory), that the laboratory technician trips over and makes it explode, or that it is simply stored incorrectly so that due to wear and tear of the insulating material over the years, the pump ends up exploding. These are all trivial cases but cannot apply to our set up.

In the logically significant ideal experiment —under controlled parameters, i.e., a laboratory with a perfect experimental setup, without failures— the *i3* interpretation is inapplicable. That is why we propose —on the principle of charity— to distinguish two definitions of photon-sensitive bomb: one *outside* the laboratory described in the thought experiment (where *i3* may be possible) and one *inside* the laboratory as described (where *i3* is simply inapplicable). Inside our laboratory, the bomb is, by definition, a device which *in fact* explodes only when the photon —coming from our light source— impacts on it. Under no other circumstances does it explode.

This answers better to a biconditional: the bomb explodes if, and only if, our photon impacts on it ($p \leftrightarrow q$) where *i3* is removed as a possibility, and all possible interpretations are listed in the following table (Table 3):

	p	q	$p \leftrightarrow q$	POSSIBILITY
i1	1	1	1	*Possible* *(The events p and q occur)*
i2	1	0	0	*Impossible*
i3	0	1	0	*Inapplicable*
i4	0	0	1	*Possible* *(The events p and q do not occur)*

Table 3. Definition of the bomb *inside* the laboratory through a biconditional.

However, we have not formally defined the counterfactual detection that takes place in IFMs yet. We must recall that the counterfactual we are looking for is that the bomb does not explode and that our photon is detected in D2. Thus, we must rule out also the option *i1*. To do so, we must include the key fact of the photon detection in D2. This leads us to the restatement of the counterfactual detection in this experiment.

We name *counterfactual detection* the combination of two complex events reflected in the following molecular propositions:

A. No photon impact on bomb, no bomb explosion;[12]

B. Photon reflection at BS2, photon detection in D2.

Only the combination of both molecular sentences expresses that a counterfactual detection of the bomb has taken place in the quantum sense. In that combination, the formula $p \leftrightarrow q$ can maybe characterize the operation of the bomb as some kind of definition. It is now necessary to formally include proposition B: "the photon reflected in BS2 impacts on D2". This is the data that will allow us to claim that there is an actual bomb inside the device —on which the photon has not impacted, and which has not exploded, i.e., *i4*. In other words, proposition B excludes possibility *i1*.

7. Classical Logic's Constraints

7.1. Counterfactual detection as a relation between two biconditionals

In accordance with the foregoing, we offer a more complete definition of counterfactual detection, as that quantum-mechanical procedure that serves to determine —given detection in D2, after reflection in BS2— the effective detection of an object that would explode when interacting with a photon, while the object does not actually explode. Consequently, the formalization of counterfactual detection must include some connection of proposition A —represented by the biconditional $p \leftrightarrow q$, in i4— with the formalization of proposition B.

Taking r as "the photon is reflected in BS2" and s as "the photon is detected in D2" we can formalize the sentence "the photon is detected in D2 if it is reflected in BS2" as another biconditional, expressing B as $r \leftrightarrow s$.

The reasons why we take the biconditional $r \leftrightarrow s$ instead of the simple conditional $r \rightarrow s$ are the same as before with $p \leftrightarrow q$, namely: we stick to an experiment performed in a laboratory with all parameters ideally controlled so that the detector D2 cannot signal the detection of a photon if this photon has not previously been reflected in BS2. And it can only be reflected in BS2 if it comes from our light source, under some ideal conditions inside the laboratory.

[12] It must be noted that this proposition contains two negations by virtue of the fact we are considering the biconditional in the interpretation *i4*.

As counterfactuality requires a connection between two biconditionals, we need some connection between them expressed as:

$$(p \leftrightarrow q)\ conn\ (r \leftrightarrow s) \tag{18}$$

which is the extended expression of the semantic relation between propositions A ($p \leftrightarrow q$ in *i4*) and B ($r \leftrightarrow s$ in *i1*), i.e. "*A conn B*".

We proceed to reformulate the IFMs flowchart in Figure 5 including the modifications discussed so far:

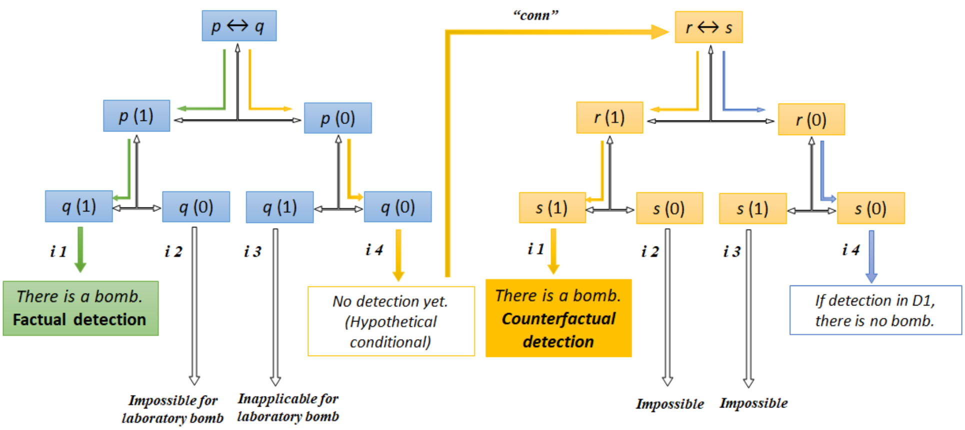

Figure 5. This figure represents diagrammatically the truth table of the biconditionals: A: $(p \leftrightarrow q)$ *and* B: $(r \leftrightarrow s)$, and their relation in the counterfactual detection process.

Next, we will elucidate what the connection between A and B may consist of.

7.2. Analyzing classical connectives as "*conn*" and their limits

In the following, we inquire into whether the counterfactual detection expressed in the terms of formula (18) can be included in a propositional language by introducing a connective to represent the relation we are looking for. We can rule out, in advance, a monadic connective (negation) which could satisfactorily formalize such a relation, since it connects two molecular propositions. In the classical propositional logic, we can define all possible connectives as possible syntactical relevant (non-repeated and well-formed) combination of truth values of each 2^2 (bivalent dyadic) combinations. We will have then that the total number of so-called connectives is 16. And we can represent all in a single truth table as in Table 4.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	B	⊤	∨	←	A	→	B	↔	∧	↑	⊻	~A	↛	~B	↚	↓	⊥
1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
1	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0
0	1	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
0	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0

Table 4: Configuration of all possible non-trivial connectives that can be defined in a bivalent classic calculus.

And from all the logical possibilities reflected in this table we want to focus only in those cases valid for those interpretations in which $v(A) = 1$ and $v(B) = 1$ and, furthermore, $(A\ conn\ B) = 1$. Let us look at the columns and see why the different connectives cannot replace "*conn*" without loss of relevant information:

(1) The first and last (1 and 16) columns must be ruled out since logical validities (⊤) and invalid formulas (⊥) cannot replace the relation we want. In classical logic, as pointed out by Wittgenstein in his *Tractatus Logico-Philosophicus* (4.461), tautologies and contradictions don't add any relevant information.

(2) The second and tenth columns are the so-called inclusive (∨) and exclusive (⊻) disjunctions, respectively.

(3) The exclusive disjunction asserts that the two formulas to be related do not bear any kind of semantical relation in their models and, therefore, could even be assumed to be equivalent to an absence of connection: *"~~conn~~".* This is because they cannot be true simultaneously under this logical operation. Inclusive disjunction, on the other hand, raises the problem of allowing only one of the two biconditionals to be true while the other is false, even though this need not necessarily be the case. Similarly, columns eleven and thirteen must be eliminated. This is because they exclude one of the two terms, equivalent to the negations of the formulas to be related, in all their interpretations. Thus, the connectives defined by columns four and six (numbers 4 and 6) that exclude the other formula, stating only the interpretation of one —that is, being logically equivalent to it— should be deleted for the same reason.

(4) The Sheffer (↑) and the Peirce (↓) strokes are removed for the same reasons relating to their monadic character, as well as for being false in the case in which both biconditionals are true —an obviously indispensable requirement. With the latter requirement, we may also discard column twelve (12) or abjunction (↛) and column fourteen (14) or converse abjunction (↚).

(5) The conditional (→), in the fifth column (5), raises the same problems reviewed in the previous section, as well as new problems. If we take "*conn*" as a

conditional, the resulting formula will be true even if, for example, both biconditionals are false —which is not an interpretation consistent with IFMs—, as well as when biconditional A corresponding to the definition of the bomb is false while B is true. Remember this was the original proposal.

(6) The converse conditional (←), third column (3), cannot be the connection because it allows (i) that A holds and B does not, and (ii) that A and B are both false, making it impossible in each case to formalize the facts involved in counterfactual detection through D2. And the biconditional (↔), seventh column (7), cannot be the connection "*conn*" because it allows A and B to be false, making it impossible, again, to formalize of the events involved in the counterfactual detection through D2.

(7) Finally, the one remaining option is the conjunction (∧) in column eight (8). In this case we will have that biconditionals A and B will only be connected when both are true —and in no other circumstances.

Let us examine the latter possibility a little more closely. We have established that, in counterfactual detection, at least four events are related:

p: "Photon impact on bomb". This event should not occur.

q: "Bomb explosion". This event should not occur.

r: "Photon reflection at BS2". This event should occur.

s: "Photon detection at D2". This event should occur.

We have defended that these events (*p* with *q* and *r* with *s*), as circumscribed to ideal laboratory conditions, have a relationship of *sufficiency* and *necessity*:

A: It is sufficient and necessary that *p* (impact of the photon on the bomb) for *q* (explosion of the bomb), and *vice versa*.

B: It is sufficient and necessary that *r* (reflection of the photon by BS2) for *s* (detection of the photon in D2) and *vice versa*.

But these relations of sufficiency and necessity can't be properly formalized through any classical connective, even the conditional as relevance paradoxes show (note that classical conditional captures a more relaxed relation). The only possible way will be to use a biconditional to ensure from right to left and *vice versa* that it is logically impossible to obtain *p* without *q* nor *r* without *s*. So we write:

A: $(p \leftrightarrow q)$

B: $(r \leftrightarrow s)$

Then, we must recall that, for counterfactual detection to occur, the events involved in A must not have occurred, while those involved in B must have occurred. This, as we have already seen, can be registered through the line of interpretation of the truth table of the biconditional, which reflects this occurrence/non-occurrence of the events:

A: $(p \leftrightarrow q)$, should be adopted by *i4*: *v(p)=0*, *v(q)=0*.

B: $(r \leftrightarrow s)$, should be adopted by *i1*: *v(r)=1*, *v(s)=1*.

Nevertheless, it remains to represent formally that, in the counterfactual detection, the two molecular events must occur: A must *arise* in *i4* and B must *arise* in *i1*, and none of the options reviewed in the previous sections, relative to other logical connectives different from the conjunction, can arise. Any other relation that arose between A and B, in any other interpretation, would correspond to an experimental situation other than counterfactual detection. Consequently, we can take the step of formalizing $(p \leftrightarrow q)\ conn\ (r \leftrightarrow s)$ as follows:

$$(p \leftrightarrow q) \wedge (r \leftrightarrow s) \tag{19}$$

However, we have not yet formally expressed all that is involved in counterfactual detection: the formula (19) does not allow us to discern whether the events involved in $p \leftrightarrow q$ and those involved in $(r \leftrightarrow s)$ occur or do not occur. In other words, in (19) we do not know which of these two alternatives obtains:

$$(p \leftrightarrow q) \wedge (r \leftrightarrow s):\ \ v(p) = 0, v(q) = 0, v(r) = 1, v(s) = 1 \tag{20 a}$$

$$(p \leftrightarrow q) \wedge (r \leftrightarrow s):\ \ v(p) = 1, v(q) = 1, v(r) = 0, v(s) = 0 \tag{20 b}$$

We should note that we have discarded the other combinations of truth values, since the experiment takes place under ideal laboratory conditions (see Table 3). And also that, in fact we are modeling here states of knowledge of quantum results.

But then we can notice that this is the same as simply asserting the truth interpretations of every propositional letter directly as

$$(p \wedge q) \wedge (\neg r \wedge \neg s) \tag{21}$$

Which is nothing but the repeated definition formalized completely *ad hoc* stipulated without any kind of generality capable to be implemented in a general framework.

Moreover, we would still have to capture the simultaneously impossibility of getting an explosion and a detection in D2. This means we need to add an exclusive disjunction between *q* and *s* until obtain:

$$\big((p \wedge q) \wedge (\neg r \wedge \neg s)\big) \wedge (q \veebar s) \tag{22}$$

And with this we can indicate that (i) Either the event q ("bomb explosion") obtains, and then only the first of the disjuncts, q, is true or else (ii) event s

("detection in D2") occurs. The latter situation corresponds to the events involved in counterfactual detection, so we can try to reduce the counterfactual detection to this case but then the logical approach collapses. The ideal of use biconditionals and affirm something like $[(p \leftrightarrow q) \wedge (r \leftrightarrow s)] \wedge (s)$ keeping the exclusive disjunction will face the interpretations in which the first biconditional is true because both p and q are false and the second one having r and s both true. Indeed, if we try to maintain the truth of s then we must interpret the second biconditional in this precisely case and this has no physical meaning here.

Classical logical connectives, therefore, exhibit a clear intrinsic limitation when attempting to model the notion of quantum counterfactuality. The main issue is that previous treatments in the literature have assumed the material conditional to be an adequate connective for approaching this concept. However, propositional logic only captures various relations between possible logical values and operations on propositions, allowing for different interpretations, such as logical possible worlds, where we intend to precisely situate them. Quantum counterfactuality, however, is qualitatively distinct from classical counterfactuality, in which one might claim, at best, that a conditional is true even when both terms it connects are false. And classical logic presents a clear limitation as an epistemological tool when approaching its modeling.

Therefore, contrary to the usual proposals we have analyzed, it seems relevant to conduct an analysis of quantum counterfactuality based on standard quantum logics. As we mentioned at the outset, these logics are nothing more than the logical interpretation of the algebra underlying the set of linearly closed subspaces of a Hilbert space, which forms a non-distributive lattice. This suggests that the shift towards quantum logics is, in all respects, a natural one. And it is precisely the partial order relation of this lattice that serves as the foundation for defining a *quantum conditional*. Moreover, the extensive literature in logic and philosophy of logic regarding the proper semantic characterization of this conditional provides us with valuable tools to address quantum counterfactuality within the Hilbert space formalism.

Additionally, (i) the fact that we can define the four atomic propositional elements necessary for the correct modeling of Interaction-Free Measurement (IFM) in terms of systems and system interactions (under operators) with four systems represented over subspaces on H, as well as (ii) the possibility of modeling the evolution of the entire system within the quantum lattice itself, seem to indicate (alongside the complete failure of classical logic to model this notion of counterfactuality) the necessity of shifting our study to quantum logics.

This shift will not only clarify the logical relations between the elements we need to consider but will also allow us to connect debates across different domains: philosophical (concerning the status of quantum logics and their relation to certain realist interpretations that, as we have seen, use quantum counterfactuality as an argument), logical (providing an opportunity to apply quantum logics to a new experiment and, in doing so, refining the notion of conditionality through counterfactuality itself), and computational (bridging the approach based on quantum logics and the Hilbert space formalism, linking algebraic logical results with quantum logic gates and the characterization of the interferometer as a quantum circuit). In the lattice of closed subspaces of a Hilbert space, propositions behave not as truth-valued sentences but as operators whose meet and join reflect the physical compatibility relations between observables. This non-distributive structure allows the formalism to represent interference, phase relations and counterfactual paths in a way that no classical connective can reproduce.

8. Concluding remarks

In this paper we have shown that the counterfactual conditional (as merely subjunctive) is insufficient to represent the quantum counterfactual detections that obtain in IFMs. We cannot formalize them using propositional logic through a connective privileging one interpretation (*i4*) over the rest. But this constraint, moreover, is not overcome by the *ad hoc* incorporation of new connectives labeled "quantum" based on a classical calculus. Either we relate two propositions through a conditional, or we do not, but limiting their connection to a concrete interpretation of the conditional implies an *a posteriori* approach. This does not yield an adequate formalism yet. Moreover, to qualify the interpretation *i4* so as to be able to transform it into another expression that represents the particular truth values of *p* and *q* as 0 returns logically equivalent (substitutable) expressions without any physical interpretation.

The quantum counterfactual detection process involves at least four items (*p*, *q*, *r*, *s*) we have successfully identified. We have studied some alternative, more complex formulas, that could formalize the relations of sufficiency and necessity existing between the pairs of facts (p, q) and (r, s), through the biconditional. But they also failed due to the excess of semantic *ad hoc* restrictions imposed. At this point we must, point out that the best classical alternatives still suffer from serious limitations. They do not represent the probabilities of occurrence of the different observables since it cannot quantify or weight the states $|boom\rangle$ and $|D2\rangle$. So, it does not capture the quantum mechanical reasons why the counterfactual detection occurs. Secondly, it may be thought to be an *ad hoc* formalization that: (i) does not generalize into a formal structure capable of accounting for the functioning of the mechanism, (ii) does not provide a comprehensive approach through the quantum-mechanical formalism involved, (iii) is not obviously computable. Our analysis indicates that classical propositional logic is insufficient —though not wholly useless— for modelling interaction-free measurements. It captures certain idealized causal dependencies but fails to represent the genuinely quantum correlations and probability structures on which the phenomenon relies. For a complete treatment, a non-Boolean framework such as standard quantum logics appears necessary.

First, in the presentation of the experiment itself, and finally, after analyzing all the limitations of classical logic, standardly defined quantum logics within the Hilbert space formalism appear to be the most promising formal tool for addressing quantum *counterfactuality*. Moreover, as we have just mentioned, one of the most critical limitations of classical logical analysis is its inability to connect with the quantum inferential processes, specifically those of quantum probability, that are necessary in this context and crucial for modeling what occurs in the bomb detector.

Thus, IFMs provide a unique testbed where physical, logical and philosophical considerations converge: the physics demands a non-Boolean structure; the logic must capture non-classical relations of dependence; and the philosophical analysis clarifies how counterfactual reasoning should be reconceived in quantum contexts. This confluence highlights the need for a unified framework and suggests that quantum logics provide precisely the level of generality required.

Quantum logics, being a natural lattice-theoretic interpretation of the algebra isomorphic to the space defined by projection operators on Hilbert space, allow for this connection in a completely natural and automatic manner. Additionally, we have also seen that quantum logics facilitate an exploration of a novel relationship between (i) philosophical debates related both to interpretations of quantum mechanics (where IFMs are used as cases to construct arguments in favor of certain many-worlds-type interpretations) and to the very concept of physical counterfactuality, (ii) discussions in the philosophy of logic concerning the formalism of quantum logic itself, and (iii) debates in computational contexts, given that quantum logics provide a natural formalism for linking the bomb detector experiment to its definition as a quantum circuit.

Nevertheless, the epistemological limitations of classical logical connectives in this context could serve as a foundation for a future critique of all those quantum-logical calculi which take propositional logic as a basis and classical connectives —especially the conjunction and the conditional— with the same truth table to generate descriptions of quantum mechanical states. What is needed is a more general formalism, one that fits the quantum interpretation of the interferometer, not as a mere matter of nomenclature, akin to a simple formal exercise in modeling sentences in natural language in propositional logic but going way beyond in fully describing the physics of the MZ interferometer. However, even though we do not yet possess it, having explored the limitations of the classical approach taken thus far allows us to explicitly identify the necessary ingredients, recognize how quantum logics emerge as the most promising framework, and highlight the next steps to be taken.[13]

References

Azuma, H. 2004. "Interaction-Free Quantum Computation." *Physical Review A* 70 (1): 012318. https://arxiv.org/abs/quant-ph/0403159.

Birkhoff, G., and J. von Neumann. 1936. "The Logic of Quantum Mechanics." *Annals of Mathematics* 37 (4): 823–843.

Collins, J., N. Hall, and L. A. Paul, eds. 2004. *Causation and Counterfactuals*. Cambridge, MA: MIT Press.

Dalla Chiara, M.L., and R. Giuntini. 2002. "Quantum Logics." In *Handbook of Philosophical Logic*, edited by D.M. Gabbay and F. Guenthner, 129–228. Dordrecht: Springer.

Elitzur, A.C., and L. Vaidman. 1993. "Quantum Mechanical Interaction-Free Measurements." *Foundations of Physics* 23: 987–997.

Engesser, K., D. M. Gabbay, and D. Lehmann, eds. 2007. *Handbook of Quantum Logic and Quantum Structures*. Amsterdam: Elsevier.

Cuesta, J.A.F., 2021. "Lógica modal megárico-estoica: posibilidad y necesidad como operadores atléticos." Human Review. International Humanities Review / *Revista Internacional de Humanidades* 10: 261–270.

Cuesta, J.A.F., and C. Sánchez Ovcharov. 2023. "Contrafácticos cuánticos: aproximación lógico-filosófica a las medidas cuánticas sin interacción." *Revista Colombiana de Filosofía de la Ciencia* 23 (46): 313–337. https://doi.org/10.18270/rcfc.v23i46.3841.

Grynberg, G., A. Aspect, and C. Fabre. 2010. *Introduction to Quantum Optics: From the Semi-Classical Approach to Quantized Light*. Cambridge: Cambridge University Press.

[13] Acknowledgment: J.A.F. Cuesta received funding from a UCM–Santander grant (CT15/23) and from the AFIM Research Project PR17/24-31887 (UCM/CAM). His research is conducted within the Mathesis and GISC research groups. Carmen Sánchez-Ovcharov and J.A.F. Cuesta also carry out their research under the national project PhilStatModNat (PID2024-159749NB-I00).

Hosten, O., M.T. Rakher, J.T. Barreiro, N.A. Peters, and P.G. Kwiat. 2006. "Counterfactual Quantum Computation Through Quantum Interrogation." *Nature* 439 (7079): 949–952.

Kwiat, P., H. Weinfurter, T. Herzog, A. Zeilinger, and M.A. Kasevich. 1995. "Interaction-Free Measurement." *Physical Review Letters* 74 (24): 4763.

Lewis, D. 1973. *Counterfactuals*. Oxford: Wiley–Blackwell.

———.1993. *On the Plurality of Worlds*. Oxford: Blackwell.

Méthot, A.A., and K. Wicker. 2001. "Interaction-Free Measurement Applied to Quantum Computation: A New CNOT Gate." arXiv:quant-ph/0109105.

Mitchison, G., and R. Jozsa. 2001. "Counterfactual Computation." *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 457 (2009): 1175–1193.

Penrose, R. 1994. *Shadows of the Mind*. Vol. 4. Oxford: Oxford University Press.

Raj, A., B. Das, B.K. Behera, and P. K. Panigrahi. 2019. "Demonstration of Bomb Detection Using the IBM Quantum Computer." Preprint in https://doi.org/10.20944/preprints201902.0232.v1.

Russo, O., and L. Jiang. 2012. "Interaction Free Measurement of Quantum Systems." In *APS March Meeting Abstracts*, Vol. 2012, K1–301.

Stalnaker, R. 1968. "A Theory of Conditionals." *Studies in Logical Theory* 2: 98–112.

Svozil, K. 1998. *Quantum Logic*. Singapore: Springer.

Vaidman, L. 1994. "On the Paradoxical Aspects of New Quantum Experiments." *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1994 (1): 211–217.

———. 2003. "The Meaning of the Interaction-Free Measurements." *Foundations of Physics* 33: 491–510.

———. 2008. *The Elitzur–Vaidman Interaction-Free Measurements*. Preprint. https://arxiv.org/abs/0801.2777.

———. 2009. "Counterfactuals in Quantum Mechanics." In *Compendium of Quantum Physics*, 132–136. Berlin: Springer.

———. 2018. "Ontology of the Wave Function and the Many-Worlds Interpretation." Preprint. http://philsci-archive.pitt.edu/14590.

———. 2019. "Analysis of Counterfactuality of Counterfactual Communication Protocols." *Physical Review A* 99 (5): 052127.

Weingartner, P. 2004. *Alternative Logics: Do Sciences Need Them?* Berlin: Springer.