NEGATIVE PROBABILITIES AND COUNTERFACTUAL REASONING ON THE DOUBLE-SLIT EXPERIMENT

J. ACACIO DE BARROS*, GARY OAS†, AND PATRICK SUPPES⋆

ABSTRACT. In this paper we attempt to establish a theory of negative (quasi) probability distributions from fundamental principles and apply it to the study of the double-slit experiment in quantum mechanics. We do so in a way that preserves the main conceptual issues intact but allow for a clearer analysis, by representing the double-slit experiment in terms of the Mach-Zehnder interferometer, and show that the main features of quantum systems relevant to the double-slit are present also in the Mach-Zehnder. This converts the problem from a continuous to a discrete random variable representation. We then show that, for the Mach-Zehnder interferometer, negative probabilities do not exist that are consistent with interference and which-path information, contrary to what Feynman believed. However, consistent with Scully et al.'s experiment, if we reduce the amount of experimental information about the system and rely on counterfactual reasoning, a joint negative probability distribution can be constructed for the Mach-Zehnder experiment.

Two of the authors (JdB, GO) would like to express their gratitude for having the honor to contribute to this volume recognizing Patrick Suppes. It is with great sadness that Pat's passing came so soon. He did not have a chance to contribute to this final version, but the core ideas put forth here stem from him, and we believe that he would be pleased with the final result. However, we emphasize that any errors are the exclusive responsibility of the first two authors, and would not be present if the paper had gone through Pat's usual rigorous review. We also want to take this opportunity to express our indebtedness to Pat for his guidance and patience over the past few decades. Pat introduced both of us to the importance of joint probability distributions in quantum mechanics, and was to us not only a collaborator and mentor, but also a friend, and we heartily dedicate this paper to him.

1. INTRODUCTION

Ever since its inception, quantum mechanics has not ceased to perplex physicists with its counter-intuitive descriptions of nature. For instance, in their famous paper Albert Einstein, Boris Podolsky, and Nathan Rosen (EPR) argued the incompleteness of the quantum mechanical description (Einstein et al., 1935). At the core of their argument was the superposition of two wavefunctions where properties of two particles far apart, A and B, were highly correlated. Since both particles are spatially separated, EPR argued that a measurement on A should not affect B. Therefore, we could use the correlation and a measurement on A to infer the value

Date: December 15, 2014. Early draft. Do not quote without authors' permission.

of a property in B *without disturbing* it. Thus, concluded EPR, the quantum mechanical description of nature had to be incomplete, as it did not allow the values of a property to be fixed before an experiment was performed.

In 1964, John Bell showed that not only quantum mechanics was incomplete, but also that a complete description of physical reality such as the one espoused by EPR was incompatible with quantum mechanical predictions (Bell, 1964). Later on, Alain Aspect, Jean Dalibard, and Gérard Roger, in an impressive and technically challenging experiment, obtained correlation measurements between measurement events separated by a spacelike interval. Their correlations supported quantum mechanics, in disagreement with EPR's metaphysical views (Aspect et al., 1981, 1982). Other puzzling results followed, like the Kochen-Specker theorem (Kochen and Specker, 1967, 1975), Wheeler's delayed choice experiment (Wheeler, 1978; Jacques et al., 2007), the quantum eraser (Scully and Druhl, 1982), and the Greenberger-Horne-Zeilinger paradox (Greenberger et al., 1989; de Barros and Suppes, 2000, 2001), to mention a few.

What most of the above examples have in common (with the exception of Kochen-Specker) is that they all use superpositions of quantum states. There is nothing more puzzling in quantum mechanics than the fact that a given system can be in a state with two incompatible properties simultaneously "present." For example, let $\hat{O}$ be an observable corresponding to a property $\mathbf{O}$, and $|o_1\rangle$ and $|o_2\rangle$ two eigenstates of $\hat{O}$ with eigenvalues o_1 and o_2. If the quantum system is in the state $|o_1\rangle$, we may say that it has an objective property $\mathbf{O}$ and its value is o_1, in the sense that if we measure this system for property $\mathbf{O}$, the outcome of this measurement will be o_1 with probability one. Similarly if the system is in the state $|o_2\rangle$. However, it is also in principle possible to prepare a system in a quantum state that is the superposition of $|o_1\rangle$ and $|o_2\rangle$, i.e. $c_1|o_1\rangle + c_2|o_2\rangle$, where c_1 and c_2 are any complex numbers satisfying the constraint $|c_1|^2 + |c_2|^2 = 1$. At a first glance, this may not seem puzzling, as it is just telling us that the system is perhaps in a state where the value of $\mathbf{O}$ is unknown, except that it can either be o_1 (o_2) with probability $|c_1|^2$ ($|c_2|^2$). The perplexing aspects of superpositions come from the study of properties such as $\mathbf{O}$ taken in conjunction with other properties, say $\mathbf{O}'$, in cases where their corresponding observables, $\hat{O}$ and $\hat{O}'$, do not commute ($[\hat{O}, \hat{O}'] \neq 0$).

There is perhaps no simpler context in which the superposition mystery reveals itself than single photon interference, realized in the double-slit experiment. In fact, in his Lectures in Physics, Richard Feynman famously claimed that this experiment contains the *only* mystery of quantum mechanics (Feynman et al., 2011). Although there are other mysteries, as Silverman (1995) pointed out, the double-slit experiment provides us with an understanding of some key aspects of quantum mechanics. It may be the true mystery of quantum mechanics lies in the idea that a property is not the same in different contexts (Dzhafarov and Kujala, 2013a; Markiewicz et al., 2013; Howard et al., 2014; de Barros et al., 2014), and that such contexts (perhaps freely) chosen by an observer can be spacelike separated (Bell, 1964, 1966; Greenberger et al., 1989; de Barros and Suppes, 2000, 2001; Dzhafarov and Kujala, 2014b).

One of the main "disturbing" mysteries of the double-slit system, as described by Feynman, is the non-monotonic character of the probabilities of detection. This was exactly what motivated Feynman to use negative probabilities to describe quantum systems (Feynman, 1987). However, as Feynman remarked, such an approach did not seem to provide any new insights into quantum mechanics.

It is our goal here to show that we can indeed gain some insight by using negative probabilities. This paper is organized the following way. First, we introduce in Section 2 a simplified version of the double-slit experiment in the form of the Mach-Zehnder interferometer. Then in Section 3, we present a theory of negative probabilities. We then show, in Section 4, that proper probability measures do not exist for the simultaneous measurements of the particle- and wave-like properties, but negative probabilities do under certain counterfactual conditions that are often assumed in experimental analyses.

2. The Mystery of the double-slit Experiment

In this section we reproduce Feynman's discussion of the double-slit experiment, and why he considered it mysterious (Feynman et al., 2011). Feynman's argument involves the idea that classically we think of systems in terms of two distinct and incompatible concepts, particles or waves[1]. Such concepts are incompatible because particles are localized and waves are not. To see this, let us start with a point particle. In classical mechanics, the main characteristic of particles is that they are objects localized in space, and therefore can only interact with other systems that are present in their localized position. For example, let us consider a particle P whose position at time t is described by the position vector $\mathbf{r}_P(t)$. At time t_0, P can only interact with another physical entity that is also at $\mathbf{r}_P(t_0)$, either another particle S such that $\mathbf{r}_S(t_0) = \mathbf{r}_P(t_0)$ or a field that is nonzero at $\mathbf{r}_P(t_0)$. For instance, when a particle is subject to no external fields (of course an idealization), such as gravity, it travels in a straight line at constant speed, since no interaction is present. If this particle then collides with another particle, say a constituent of a wall placed in the way of the original particle, an interaction will appear[2]. But, as soon as the particle looses contact with the wall, the interaction ceases to exist. In other words, particles interact locally.

The second basic concept is that of waves. Historically, the physics describing a point particle was extended to include the description of continuous media, and, more important to our current discussion, the vibrations of such media in the form of waves. Waves, therefore, were considered vibrations of a medium made out of several point particles, and the local interactions between two neighboring particles would allow for a perturbation in one point of the medium to be propagated to another point of the medium. Without going into the discussion of the particulars of electromagnetic waves, the main point is that because a single particle has an infinite number of neighbors, a disturbance on its position propagates to *all* of its neighbors and to *all* directions. Thus, the effect of such perturbation on particle S' belonging to this medium due to a perturbation on particle P' does not depend on the direct contact of S' and P'. More importantly, such effect depends not only on P', but also possibly on all other particles that make up the medium, and also on all interactions or boundary conditions that such particles need to satisfy. In other words, waves interact non-locally.

[1] You could also have fields, but in the context of the double-slit experiment, as it will become clear later, the relevant property of a field would be its spatial oscillations as a wave.

[2] There is, of course, the obvious issue of how could this interaction be relevant, given that it would occur with probability zero.

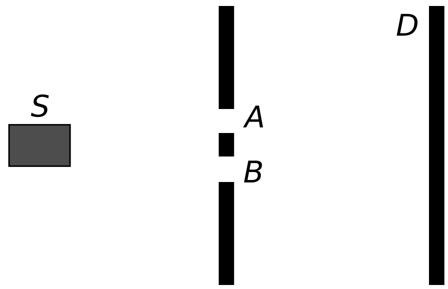

FIGURE 2.1. Schematics of the double-slit experiment. A source S emits a physical object towards a barrier, where two slits, A and B, are cut to allow for its passage. Then, at a screen D, the object is detected.

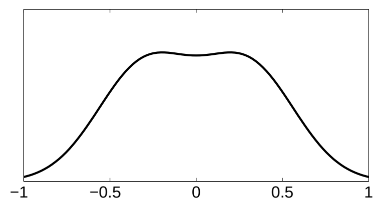

FIGURE 2.2. Probability density of observation for the double-slit experiment, assuming a particle model.

Going back to the double-slit experiment, let us analyze it from the point of view of what one should expect to happen were it being modeled with particles or with waves. Figure 2.1 shows a typical double-slit setup. We start with particles. Let us assume that S sends particles in random directions. A particle leaving S would interact only locally with the barrier and nothing else. This means that in between S and the barrier, this particle travels along a straight line. Once it reaches the barrier, it either goes through one of the slits and reaches D, or is reflected back (or absorbed, depending on the barrier's properties). While going through a slit, the particle may interact with the walls, perhaps bouncing off of it, and therefore causing some scattering from the direct path between S and D. Thus, if we run this experiment many times, we should expect the observed probability distribution of particles on D to be somewhat like as depicted in Figure 2.2. The resulting probability is simply the (normalized) sum of the probability of a particle going through slit A and B.

A wave analysis of the experiment shows something quite different. First, a wave is the result of a perturbation of a medium. In this case, the source S disturbs the medium, and such disturbance is propagated in all directions. One characteristic of such propagation is that its speed is dependent on the medium, and for the double slit experiment this is reflected in the arrival of a wave crest (or valley) in A at the same time that a crest (or valley) arrives in B. If A and B are small compared to the wavelengths, we can think of them as secondary wave sources oscillating in phase. Thus, when they arrive at D, in some places they will be in phase, whereas in other

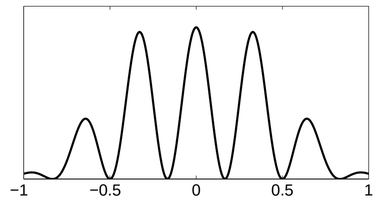

FIGURE 2.3. Intensity (arbitrary units) of the wave at D for the double-slit experiment.

places they will be out of phase. The result is the constructive and destructive interference pattern that we know for waves, shown in Figure 2.3.

Now for the puzzling aspect of it all: quantum systems, e.g. electrons, have particle-like characteristics while at the same time being prone to wave-like interference. Here is how those aspects manifest themselves. Experimentally, electrons are particles. We know this because they are localized, in the sense that when we measure an electron, it shows up as a point in a fluorescent screen or a localized detector. This is to be contrasted with waves, which are spread out, and therefore have measurable components (i.e., momentum and energy) in more than one place. Thus, as particles, when an electron leaves the source S, it will either go to slit A or B. Since its interactions are solely local, if it goes through, say, A, it can only interact with A, and not with B. Therefore, to a particle, it makes no difference at all if slit B is open or not when it goes through A, and *vice versa*. This is why we should expect a distribution like that on Figure 2.2. The disturbing fact is that an electron, if both slits are open, satisfies, after many runs, the distribution given in Figure 2.3. So, here is the puzzle. For a wave, the intensity is zero at several points, e.g. at position 0.5. How can this zero intensity be understood? How can the particle "know" (in the sense of interacting) about B if only interacting locally with A? To make this point even clearer, we will examine this question in detail in Section 4, using a simplified case of the double-slit experiment, the Mach-Zehnder interferometer, together with a framework of extended probabilities. But first, let us examine the concept of negative probabilities.

3. NEGATIVE PROBABILITIES

In this Section we lay out the main relevant results and definitions for negative probabilities. We start by first defining it in a way that is related to Kolmogorov's (1950) probability measure. We then prove some simple but relevant results.

Definition 1. Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, and p a real-valued functions, $p : \mathcal{F} \to \mathbb{R}$. Then $(\Omega, \mathcal{F}, p)$ is a probability space, and p a probability measure, if and only if:

$$\begin{aligned}
&\text{K1.} \quad && 0 \le p\left(\{\omega_i\}\right) \le 1, \quad \forall \omega_i \in \Omega \\
&\text{K2.} \quad && p\left(\Omega\right) = 1, \\
&\text{K3.} \quad && p\left(\{\omega_i, \omega_j\}\right) = p\left(\{\omega_i\}\right) + p\left(\{\omega_j\}\right), \quad i \ne j.
\end{aligned}$$

The elements ω_i of Ω are called *elementary probability events* or simply *elementary events*.

Definition 1 is the finite version of Kolmogorov's standard definition of probability measure. In the usual definition, Ω can be an infinite set, and then $\mathcal{F}$ needs to be a σ-algebra. However, for the purpose of this article, we will restrict our discussions to finite sets Ω. For that reason, we will also refer to p as a *proper probability distribution or joint probability distribution.*

It is a well know fact that for some systems it is not possible to define a proper probability distribution. This is, in fact, the heart of Bell's inequalities showing that for an EPR-Bohm type experiment no local hidden-variable theories exist that are compatible with quantum mechanics (Bell, 1964, 1966). In fact, a hidden variable exists if and only if a joint probability distribution exists (Suppes and Zanotti, 1981; Fine, 1982), and Bell's inequalities are a necessary and sufficient condition for the existence of a joint probability distribution (Fine, 1982; Suppes et al., 1996a).

To overcome this difficulty, the use of upper probabilities, where K3 is modified to include subadditivity, has been proposed (Suppes and Zanotti, 1991; de Barros and Suppes, 2001; Hartmann and Suppes, 2010). The main reason for such a proposal is that quantum mechanics, as suggested by Feynman's remarks, is nonmonotonic, and upper probabilities offer a framework where nonmonotonicity can be described mathematically. Thus, before we define negative probabilities, it is useful to start with the more well-known theory of upper probabilities.

Definition 2. Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, and p^* a real-valued functions, $p^* : \mathcal{F} \to \mathbb{R}$. Then $(\Omega, \mathcal{F}, p^*)$ is an upper-probability space, and p^* an upper-probability measure, if and only if:

$$\begin{array}{ll} \text{U1.} & 0 \leq p^*\left(\{\omega_i\}\right) \leq 1, \quad \forall \omega_i \in \Omega \\ \text{U2.} & p^*\left(\Omega\right) = 1, \\ \text{U3.} & p^*\left(\{\omega_i, \omega_j\}\right) \leq p^*\left(\{\omega_i\}\right) + p^*\left(\{\omega_j\}\right), \quad i \neq j. \end{array}$$

Remark 3. The main difference between upper and proper probabilities is the substitution K3 for U3.

Remark 4. In many systems of interest, where the probabilities of elementary events are computed from a set of given marginal probabilities, the inequalities from U3 imply an underdetermination for the possible values of the joint upper probability distribution for all $\omega_i \in \Omega$.

Remark 5. If follows from K2 and K3 that

$$\sum_{\omega_i \in \Omega} p\left(\{\omega_i\}\right) = 1,$$

but because of U3 is

$$\sum_{\omega_i \in \Omega} p^*\left(\{\omega_i\}\right) \geq 1.$$

One of the main difficulties with upper probabilities is that, because it uses subadditivity, it is very hard in practice to compute it. Subaditivity also implies that a large number of different upper measures exist, even when all moments are given. An usual approach is to request p^* to be as close as possible to a proper measure by minimizing the value of $\sum_{\omega_i \in \Omega} p^*\left(\{\omega_i\}\right)$, which can be greater than one when no proper joint exists (Suppes and Zanotti, 1981, 1991; Fine, 1994; de Barros and Suppes, 2001; Hartmann and Suppes, 2010).

Another possible approach was proposed by Feynman (1987) in connection to the two slit experiment: negative probabilities. Though Feynman could not find any use for negative probability, recent research has shown that there may be some advantage for using them (Abramsky and Brandenburger, 2011; Al-Safi and Short, 2013; Zhu et al., 2013; de Barros, 2014; Oas et al., 2014; Abramsky and Brandenburger, 2014; de Barros and Oas, 2014; de Barros et al., 2014). To define negative probabilities, we first need to set forth a description of certain systems where no proper probability distribution exists. This is the goal of the following definition.

Definition 6. Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, and let $(\Omega_i, \mathcal{F}_i, p_i)$, $i = 1, \ldots, n$, a set of n probability spaces, $\mathcal{F}_i \subseteq \mathcal{F}$ and $\Omega_i \subseteq \Omega$. Then $(\Omega, \mathcal{F}, p)$,where p is a real-valued function, $p : \mathcal{F} \to [0, 1]$, $p(\Omega) = 1$, is *compatible* with the probabilities p_i's if and only if

$$\forall \left(x \in \mathcal{F}_i\right) \left(p_i\left(x\right) = p\left(x\right)\right).$$

Furthermore, the marginals p_i are *viable* if and only if p is a probability measure.

Remark 7. Intuitively, we can think of the p_i's as observable marginals. The definition above says that such marginals are *viable* if it is possible to sew them together to produce a larger probability function over the whole space Ω (in the same spirit of Dzhafarov and Kujala (2013a, 2014d); de Barros et al. (2014)). Our definition is an extension of Halliwell and Yearsley (2013), as we consider not only the case where viable distributions exist, but also when they do not.

For some experimental situations, such as the EPR-Bell setup, the marginals are not viable, but are compatible with a p that has the characteristic of being negative for some elements of Ω (but not negative for the observable marginals). This motivates the following definition of a p that may take negative values.

Definition 8. Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, P and P' real-valued functions, $P : \mathcal{F} \to \mathbb{R}$, $P' : \mathcal{F} \to \mathbb{R}$, and let $(\Omega_i, \mathcal{F}_i, p_i)$, $i = 1, \ldots, n$, a set of n probability spaces, $\mathcal{F}_i \subset \mathcal{F}$ and $\Omega_i \subseteq \Omega$. Then $(\Omega, \mathcal{F}, P)$ is a negative probability space, and P a negative probability, if and only if $(\Omega, \mathcal{F}, P)$ is compatible with the probabilities p_i's and

$$\text{N1.}\qquad \forall \left(P'\right) \left(\sum_{\omega_i \in \Omega} \left|P\left(\{\omega_i\}\right)\right| \leq \sum_{\omega_i \in \Omega} \left|P'\left(\{\omega_i\}\right)\right|\right)$$

$$\text{N2.}\qquad \sum_{\omega_i \in \Omega} P\left(\{\omega_i\}\right) = 1$$

$$\text{N3.}\qquad P\left(\{\omega_i, \omega_j\}\right) = P\left(\{\omega_i\}\right) + P\left(\{\omega_j\}\right), \quad i \neq j.$$

In the above definition, we replace axiom K1 of nonnegativity with a minimization of the L1 norm of the function P. Intuitively, as with uppers, we seek a quasi-probability distribution that is as close to a proper distribution as possible. Furthermore, the departure from such proper distributions, which would have no negative numbers, motivates the following definition of M^* as a measure of this departure. Throughout this paper we use p for proper probability measures (Definition 1), p^* for upper and lower probabilities (Definition 2), and P for negative probabilities (Definition 8).

Definition 9. Let $(\Omega, \mathcal{F}, P)$ be a negative probability space. Then, the *minimum L1 probability norm*, denoted M^*, or simply *minimum probability norm*, is given by $M^* = \sum_{\omega_i \in \Omega} |P(\{\omega_i\})|$.

Proposition 10. *Let $(\Omega, \mathcal{F}, P)$ be a negative probability space and $(\Omega, \mathcal{F}, p)$ a (Kolmogorov) probability space. Then $p = P$ iff $M^* = \sum p_i$.*

Proof. Let us start with $M^* = \sum p_i$. It follows from it that all elementary events satisfy the condition $0 \leq P(\{\omega_i\}) \leq 1$, which is K1. Together with N2 and N3, then P is also a probability measure, and $(\Omega, \mathcal{F}, P)$ a probability space. Now, if $(\Omega, \mathcal{F}, p)$ is a probability space, it follows from K1 that $M^* = \sum_{\omega_i \in \Omega} |P(\{\omega_i\})| = \sum p_i$. $\square$

Remark 11. Proposition 10 tells us that axioms N1-N3 include, as a special case, K1-K3. In other words, in the special case when a proper Kolmogorovian distribution exists ($M^* = \sum p_i$), P coincides with p.

We end this section with one last definition that is relevant to physical systems.

Definition 12. Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, and let $(\Omega_i, \mathcal{F}_i, p_i)$, $i = 1, \ldots, n$, a collection of n probability spaces, $\mathcal{F}_i \subseteq \mathcal{F}$ and $\Omega_i \subseteq \Omega$. Then the probabilities p_i are *contextually biased*[3] if there exists an a in $\mathcal{F}_i$ and in $\mathcal{F}_j$, $i \neq j$, $b \neq a \neq b'$, $\sum_{\forall b \in \mathcal{F}_j} p(a \cap b) \neq \sum_{\forall b' \in \mathcal{F}_i} p(a \cap b')$.

Remark 13. In physics, for multipartite systems, this definition is equivalent to the no-signaling condition.

Proposition 14. *Let Ω be a finite set, $\mathcal{F}$ an algebra over Ω, and let $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, \ldots, n$, a set of n probability spaces, $\mathcal{F}_i \subseteq \mathcal{F}$ and $\Omega_i \subseteq \Omega$. The probabilities P_i are not* contextuality biased *if and only if there exists a negative probability $(\Omega, \mathcal{F}, P)$ compatible with the p_i's.*

Proof. See Al-Safi and Short (2013); Oas et al. (2014); Abramsky and Brandenburger (2014) for different proofs. $\square$

3.1. **Interpretations of Negative Probabilities.** As we mentioned above, both Dirac and Feynman saw negative probabilities as computational devices. Though we can take such pragmatic view, as negative probabilities help explore certain situations of interest in quantum mechanics (see Oas et al. (2014) for an example), the question still remains as to their meaning. Here we discuss some proposals on how to interpret negative probabilities.

Let us start with the interpretation of negative probabilities in terms of two disjoint probability measures, μ^+ and μ^-, initially suggested by Burgin (2010); Burgin and Meissner (2012) and then expanded and formalized by Abramsky and Brandenburger (2014). Here we follow the interpretation as presented by Abramsky and Brandenburger (2014). The main idea of this interpretation comes from the well-known fact (see Rao and Rao (1983)) that it is possible to decompose a signed measure μ into two non-negative ones, μ^- and μ^+, such that

$$\mu = \mu^+ - \mu^-.$$

Following this idea, Abramsky and Brandenburger (2014) creates two copies of the sample space, namely $\Omega \times \{+, -\}$, giving them a new dimension corresponding to $+$ or $-$. For example, for the case of three random variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, the set of all

[3] Here we adopt and adapt the terminology of Dzhafarov and Kujala (2014d).

elementary events would be $\{\omega_{xyz}, \omega_{xy\overline{z}}, \ldots, \omega_{\overline{xy}z}, \omega_{\overline{xyz}}\}$, whereas the expanded set $\Omega \times \{+, -\}$ would have as elementary events $\{\omega_{xyz+}, \ldots, \omega_{\overline{xyz}+}, \omega_{xyz-}, \ldots, \omega_{\overline{xyz}-}\}$. Because of the above decomposition, they define a probability over the set $+$ and $-$ such that

$$P = p^{+} - p^{-},$$

where now P is the negative probability and p^{+} and p^{-} can be interpreted as proper probability distributions over $(\Omega, \pm)$. To interpret P, Abramsky and Brandenburger (2014) proposes an effect akin to interference. When we observe an event, say corresponding to the element ω_{xyz+}, we use p^{+} as the distribution to create our data table, and similarly for ω_{xyz-}. However, the counting due to a $-$ element can annihilate a counting for a $+$ element, and vice versa. In a certain sense, this interpretation of negative probabilities is conceptually similar to some hidden variable approaches in the literature, as for example the virtual photon model of Suppes and de Barros (Suppes et al., 1996c; Suppes and de Barros, 1994b; Suppes et al., 1996b,d; Suppes and de Barros, 1994a; Suppes et al., 1996a) or the ESR model (Garola and Sozzo, 2009; Sozzo and Garola, 2010; Garola and Sozzo, 2011a,b; Garola et al., 2014), to cite a few. In these approaches, an underlying hidden process can erase an outcome that would be possible if it were not for the interference of non-observable events. However, the problem with this interpretation is that, even though it is based on a frequentist view, it does not provide a way of counting actual observable clicks on a measurement device and interpret them as negative probabilities; in other words, it assumes some non-accessible reality.

We now turn to another frequentist interpretation of negative probabilities, this one proposed by Khrennikov (Khrennikov, 1993a,b, 1994b,a, 2009). Khrennikov starts with the idea that, in the frequentist view, the probability of an event is defined as by the number of times such an event occurs in an infinite sequence of possible outcomes or ensembles. Following Khrennikov (2009), let S_N be a sequence of N ensembles with, $S_N = \{s_1, s_2, \ldots, s_N\}$. For each of the ensembles s_i, one can ask whether the property represented by the random variable $\mathbf{A}$ has the value a or not, and let $S(\mathbf{A} = a)$ be the subset of all ensembles such that $\mathbf{A} = a$. Then, in the standard frequentist interpretation, the probability $p(\mathbf{A} = a)$ is given by

$$p(\mathbf{A} = a) = \lim_{N \to \infty} \frac{|S(\mathbf{A} = a) \cap S_N|}{|S|}, \tag{3.1}$$

where $|\cdot|$ represents the cardinality of a set. Khrennikov then argues that there are ensembles for which the limit in (3.1) does not converge, and for such cases negative probabilities can be obtained as the result of a regularization procedure or order. In such sense, negative probabilities come as the result of quasi-random sequences that violate the principle of statistical stabilization (Khrennikov, 1993a). Khrennikov then proposes to generalize probabilities coming from sequences that violate the principle of statistical stabilization as measures taking values not only on the field $\mathbb{R}$, but also on the p-adic extensions of the set of rationals $\mathbb{Q}$, i.e. $\mathbb{Q}_p$. We recall that $\mathbb{R}$ is defined, through Cauchy sequences, as the completion of $\mathbb{Q}$ under the Euclidian norm. Similarly, $\mathbb{Q}_p$ is the completion of $\mathbb{Q}$ under the p-adic norm (see Khrennikov (2009) for a clear exposition of $\mathbb{Q}_p$ and its properties). Once he does that, he shows that certain sequences that have probability zero in the sense of (3.1) would have negative probabilities in their p-adic extension, whereas sequences of probability one would have p-adic values greater than one. Thus, according to Khrennikov, we can interpret negative (greater than one) probabilities as events of probability zero (one) for sequences that violate the principle of statistical stabilization. Here we note that contextual random variables are quasi-random, and violate the principle of statistical stabilization.

We now turn to the meaning of the minimum L1 norm we propose for negative probabilities. Similarly to negative probabilities, the sub or super additivity of upper and lower probabilities allows for a large number of solutions to the joint probability that is consistent with the marginals. One possibility is to think of upper and lowers as subjective measures of belief based on inconsistent information (Suppes and Zanotti, 1991). As such, it can be argued that, since upper and lowers do not add to one as standard probabilities do, one should choose among the many different distributions those whose sums are as close to one as possible. This is, in a certain sense, similar to what the minimum L1 norm does for negative probabilities. As such, this norm, which quantifies how much a negative probability deviates from a proper probability, provides us a measure of how inconsistent the correlations between random variables are (de Barros, 2014).

We end this section with one last general comment. Instead of using negative probabilities, it is possible to simply extend the probability space such that when we talk about correlations between experimentally observable variables, as proposed by Dzhafarov and Kujala (2012, 2013a). To understand this point, imagine we start with three variables $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$, as in the above example. Instead of thinking of them as three variables, we could think of them as six, one for each experimental context: $\mathbf{X}_Y$, $\mathbf{X}_Z$, ..., $\mathbf{Z}_Y$. It is easy to show that in some important physical examples, such as the famous Bell-EPR setup, such extension of the probability space is sufficient to grant the existence of a joint probability distribution, but at the cost of having $\mathbf{X}_Y \neq \mathbf{X}_Z$. Thus, the apparent inconsistencies mentioned in the previous paragraph could be argued to come from an identity assumption for the random variables: that a random variable remains the same in different contexts (Dzhafarov and Kujala, 2013b,a, 2014d,c). As such, the minimum norm could be interpreted as a measure of contextuality (de Barros et al., 2014; de Barros and Oas, 2014; Oas et al., 2014).

4. The Mach-Zehnder Interferometer and Negative Probabilities

Now that we saw the basic relationships between negative probabilities and upper and Kolmogorovian probabilities, we turn our attention to the two slit experiment its simplified version of the Mach-Zehnder interferometer, schematically shown in Figure 4.1. We will not attempt to give a full quantum-mechanical description of this experiment, but instead focus on an elementary representation of the main ideas behind it. Intuitively, each arm of the interferometer, A and B, corresponds to a possible path the particle can take, which is equivalent to a particle going through one slit or the other on the two slit experiment. However, the Mach-Zehnder differs from the double-slit experiment as the particle only has two possible outcomes, a detection in D_1 or D_2, as opposed to an infinite number of points on a screen. Such two possible outcomes could be thought as two points on the screen on the screen corresponding to a maximum and minimum intensities in the interference pattern.

To elaborate on the analogy with the two slit experiment, let us think about the experiment in terms of particles. First, we can select an interferometer setup such that we have constructive interference in D_1 and destructive in D_2. In other

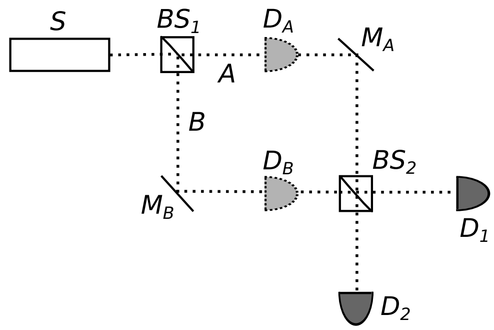

FIGURE 4.1. Schematics of a Mach-Zehnder interferometer. A light beam from a source S is divided into two equal-intensity beams by the beamsplitter BS_1. The beams are reflected by mirrors M_A and M_B, and then recombined by the beamsplitter BS_2. Photons from S are then detected by photodetectors D_1 or D_2. The count rates on D_1 and D_2 depend on the geometry of the system, in particular the optical distances between BS_1, BS_2, and M_A and M_B. In the which-path version of this experiment, detectors D_A and/or D_B may be placed on each arm of the interferometer to determine the trajectory of the photon.

words, the lengths of the interferometer arms are chosen such that $P(\mathbf{D}_1 = 1) = 1$ and $P(\mathbf{D}_2 = 1) = 0$, where $\mathbf{D}_1$ ($\mathbf{D}_2$) is a ± 1-valued random variable representing a detection on D_1 (D_2) when its value is 1 and no detection when -1. From now on we will use the standard notation $p_{d_1} = P(\mathbf{D}_1 = 1)$, $p_{\overline{d}_1} = P(\mathbf{D}_1 = -1)$, $p_{\overline{d}_2} = P(\mathbf{D}_2 = -1)$ and so on. With this notation, our interferometer is such that $p_{d_1} = p_{\overline{d}_2} = 1$ and $p_{\overline{d}_1} = p_{d_2} = 0$.

Now that the interferometer is set up, let us examine the two possible classical models (according to Feynman) behind it: the wave and particle models. We start with the wave point of view. Let $\psi = A\cos(\omega t)$ represent a coherent wave arriving at the beam splitter BS_1 at time t and being split in both directions, A *and* B. The wave going through A is unchanged by BS_1, and arrives at M_A as $\frac{A}{2}\cos(\omega t + \phi_1)$, where ϕ_1 is a phase that depends on the geometry of the interferometer, specifically on the distance between BS_1 and M_A. At M_A it becomes $-\frac{A}{2}\sin(\omega t + \phi_1)$ due to a $\pi/2$ phase shift upon reflection, and arrives at BS_2 as $-\frac{A}{2}\sin(\omega t + \phi_1 + \phi_2)$. For the wave going through B, it arrives at M_B as $-\frac{A}{2}\sin(\omega t + \phi_2)$ and at BS_2 as $-\frac{A}{2}\cos(\omega t + \phi_2 + \phi_1)$, where we assume for the geometry that distance between BS_1 and M_B is the same as the distance between M_A and BS_2 (and similarly for $BS_1$to M_A and M_B to BS_2). The beam splitter BS_2 now recombines the two waves coming from A and B, and the outputs on D_1 and D_2 are the superposition of those waves. In other words,

$$
\begin{aligned}
(4.1) \qquad \psi_{D_1} &= -\frac{A}{2}\sin\left(\omega t + \phi_1 + \phi_2 + \frac{\pi}{2}\right) - \frac{A}{2}\cos(\omega t + \phi_2 + \phi_1) \\
&= -\frac{A}{2}\cos(\omega t + \phi) - \frac{A}{2}\cos(\omega t + \phi) \\
&= -A\cos(\omega t + \phi),
\end{aligned}
$$

where the first term on the rhs is the reflected wave from A, and $\phi = \phi_1 + \phi_2$. For D_2 we obtain, with now the wave B getting a phase of $\pi/2$,

$$\psi_{D_2} = -\frac{A}{2}\sin(\omega t + \phi) - \frac{A}{2}\cos\left(\omega t + \phi + \frac{\pi}{2}\right) \tag{4.2}$$

$$= -\frac{A}{2}\sin(\omega t + \phi) + \frac{A}{2}\sin(\omega t + \phi) = 0. \tag{4.3}$$

We can now compute the mean intensity of the entering wave, ψ,

$$I_S = \langle \psi^2 \rangle_t = \frac{A^2}{2},$$

where

$$\langle f \rangle_t = \frac{1}{T}\lim_{\omega T \gg 1}\int_t^{t+T} f\, dt'$$

represents the time average. The intensity at D_1, is

$$I_{D_1} = \frac{A^2}{2},$$

whereas the intensity at D_2 is

$$I_{D_2} = 0,$$

consistent with the value for the source S. These particular values for the intensities at D_1 and D_2 present the highest contrast between the intensities at each detector, or, as it is often referred, maximum visibility (of interference). So, to summarize, according to the wave model we see no wave energy arriving at D_2 because of destructive interference due to the relative phases of the different paths the wave traveled.

Now let us examine the view that photons are particles, and let us assume that we can control the intensity of the source such that one particle at a time goes through the interferometer. A particle comes out of the source S and enters the interferometer through the beam splitter BS_1. Beam splitters divide beams into two equal intensity ones. This translates into a particle having probability 1/2 of going to either arm A or B. For the sake of argument, let us assume that the particle went into arm A. Once it leaves the interferometer, it is reflected at mirror M_A and reaches another beam splitter BS_2. Once again, it has probability 1/2 of going on either direction, since it interacts only locally with BS_2. In other words, it cannot possibly have any information about the geometry of path B, or even if it is not simply closed with the presence of a physical barrier. Therefore, the probability of this particle reaching D_1 is the same as D_2, and it equals 1/2. The same analysis can be applied to the photon going through arm B. Therefore, from a particle point of view, $p_{d_1} = p_{d_2} = 1/2$. This is in stark contradiction with the wave result.

In the standard interpretation of quantum mechanics, this contradiction is resolved by stating that one cannot simultaneously assign two complimentary properties to a quantum system. In the above case, we cannot assign the property of going through path A or B (which is what happens if we have a particle). To be able to say that a particle went through A or B, we need to actually place a detector D_A and D_B in the paths. At the same time, if we place such detectors, we destroy the wave-like behavior, and its associated probabilities at D_1 and D_2. If the detectors simply destroy the particle, then we have obviously an impossibility in obtaining the joint probability distribution in a trivial way, as we can show by the following simple example (which we also spell out in more detail below).

Let D_X and D_Y be two detectors that absorb photons, and let us put D_Y at the end of a source S that produces photons. So, if a photon is emitted, the probability of observing it is $p_{d_y} = 1$. However, if we put D_X in between S and D_Y, we will have that $p_{d_x} = 1$ and $p_{d_y} = 0$. The observable terms are the following.

$$p_{d_x d_y} = p_{\overline{d}_x \overline{d}_y} = p_{\overline{d}_x d_y} = 0,$$

$$p_{d_x \overline{d}_y} = 1.$$

But this leads to a contradiction, as

$$p_{d_y} = 1 = p_{d_x d_y} + p_{\overline{d}_x d_y} = 0.$$

That contradiction comes from an obvious reason: we have different experiments, and therefore the random variable $\mathbf{D}_Y$ representing a measurement in one experiment cannot be the same as the $\mathbf{D}_Y$ in the other experiment. The assumption of the existence of a joint distribution is equivalent to the assumption that both $\mathbf{D}_Y$'s are the same.

In the Mach-Zehnder, an analogous case to the example above would be the following. With the setup in Figure 4.1, we split the experiment into two types: destructive and non-destructive measurements. A destructive measurement happens when the observed system is not available for any other measurements afterward. For example, in many photodetection apparatuses, the photon is absorbed by the device and ceases to exist. A non-destructive measurement is the one where the system is available for later measurements. For each type of experiment, there are four possible experimental conditions, which we label as Case 1 to Case 8. We start with destructive measurements.

Case 1 (D_1, D_2 only): This case corresponds to the standard Mach-Zehnder with no *which-path* information, since no detector is put on either arm of the interferometer. Thus, a joint probability distribution exists for all the random variables involved. When this is the case, we have that $p_{d_1\overline{d}_2} = 1$ and $p_{d_1 d_2} = p_{\overline{d}_1 d_2} = p_{\overline{d}_1\overline{d}_2} = 0$. Here $p_{d_1 d_2}$ and $p_{\overline{d}_1\overline{d}_2}$ are set to zero almost by definition, as we are considering cases where we have one and only one photo-detection.

Case 2 (D_1, D_2, D_A): In this case, if we have a detection on D_A, we have no detection on D_1 and D_2 (intuitively, the photon was absorbed by the detector). On the other hand, if we have no detection on D_A, D_1 and D_2 are equiprobable, since the interference effects are destroyed by the presence of a detection. Thus, $p_{d_a d_1 d_2} = p_{d_a d_1 \overline{d}_2} = p_{d_a \overline{d}_1 \overline{d}_2} = p_{d_a \overline{d}_1 d_2} = p_{\overline{d}_a d_1 d_2} = p_{\overline{d}_a \overline{d}_1 \overline{d}_2} = 0$, and $p_{\overline{d}_a d_1 \overline{d}_2} = p_{\overline{d}_a \overline{d}_1 d_2} = \frac{1}{2}$.

Case 3 (D_1, D_2, D_B): Similarly to Case 2 above, here $p_{d_b d_1 d_2} = p_{d_b d_1 \overline{d}_2} = p_{d_b \overline{d}_1 \overline{d}_2} = p_{d_b \overline{d}_1 d_2} = p_{\overline{d}_b d_1 d_2} = p_{\overline{d}_b \overline{d}_1 \overline{d}_2} = 0$, and $p_{\overline{d}_b d_1 \overline{d}_2} = p_{\overline{d}_b \overline{d}_1 d_2} = \frac{1}{2}$.

Case 4 (D_1, D_2, D_A, and D_B): This simply tells us that we can only observe in A or B, but nowhere else, since the detectors in A and B destroy the photon, not allowing it to reach D_1 or D_2. Then $p_{d_a d_b d_1 d_2} = p_{d_a d_b \overline{d}_1 d_2} = p_{d_a d_b d_1 \overline{d}_2} = p_{d_a d_b \overline{d}_1 \overline{d}_2} = 0$, $p_{\overline{d}_a d_b d_1 d_2} = p_{\overline{d}_a d_b \overline{d}_1 d_2} = p_{\overline{d}_a d_b d_1 \overline{d}_2} = 0$, $p_{d_a \overline{d}_b d_1 d_2} = p_{d_a \overline{d}_b \overline{d}_1 d_2} = p_{d_a \overline{d}_b d_1 \overline{d}_2} = 0$, $p_{\overline{d}_a \overline{d}_b d_1 d_2} = p_{\overline{d}_a \overline{d}_b \overline{d}_1 d_2} = p_{\overline{d}_a \overline{d}_b d_1 \overline{d}_2} = p_{\overline{d}_a \overline{d}_b \overline{d}_1 \overline{d}_2} = 0$, and $p_{\overline{d}_a d_b \overline{d}_1 \overline{d}_2} = p_{d_a \overline{d}_b \overline{d}_1 \overline{d}_2} = \frac{1}{2}$.

It is easy to see that we have inconsistencies between the random variables for each case, because Case 4 gives us a joint probability distribution for all observables that is inconsistent with Case 1. This simply tells us that each experimental context gives different distributions to the random variables, as the marginal expectations are different for each experimental condition.

For the non-destructive measurement we have the following experimental outcomes.

Case 5 (D_1, D_2 only): This is clearly identical to Case 1, where $p_{d_1\overline{d}_2} = 1$ and $p_{d_1 d_2} = p_{\overline{d}_1 d_2} = p_{\overline{d}_1\overline{d}_2} = 0$.

Case 6 (D_1, D_2, D_A): In this case, if we have a detection on D_A, but there will also be a detection on either D_1 or D_2. Furthermore, regardless of the outcomes on D_A, detections on D_1 and D_2 are equiprobable, since the interference effects are destroyed by the presence of a detection. Thus, $p_{d_a d_1 d_2} = p_{d_a\overline{d}_1\overline{d}_2} = p_{\overline{d}_a d_1 d_2} = p_{\overline{d}_a\overline{d}_1\overline{d}_2} = 0$, and $p_{d_a d_1\overline{d}_2} = p_{d_a\overline{d}_1 d_2} = p_{\overline{d}_a d_1\overline{d}_2} = p_{\overline{d}_a\overline{d}_1 d_2} = \frac{1}{4}$.

Case 7 (D_1, D_2, D_B): Similarly to Case 2 above, here $p_{d_b d_1 d_2} = p_{d_b\overline{d}_1\overline{d}_2} = p_{\overline{d}_b d_1 d_2} = p_{\overline{d}_b\overline{d}_1\overline{d}_2} = 0$, and $p_{d_b d_1\overline{d}_2} = p_{d_b\overline{d}_1 d_2} = p_{\overline{d}_b d_1\overline{d}_2} = p_{\overline{d}_b\overline{d}_1 d_2} = \frac{1}{4}$.

Case 8 (D_1, D_2, D_A, and D_B): This simply tells us that we can only observe in A or B, but nowhere else, since the detectors in A and B absorb the photon, not letting it reach D_1 or D_2. Then $p_{d_a d_b d_1 d_2} = p_{d_a d_b\overline{d}_1 d_2} = p_{d_a d_b d_1\overline{d}_2} = p_{d_a d_b\overline{d}_1\overline{d}_2} = p_{\overline{d}_a d_b d_1 d_2} = p_{d_a\overline{d}_b d_1 d_2} = p_{\overline{d}_a\overline{d}_b d_1 d_2} = p_{\overline{d}_a\overline{d}_b\overline{d}_1 d_2} = p_{\overline{d}_a\overline{d}_b d_1\overline{d}_2} = p_{\overline{d}_a\overline{d}_b\overline{d}_1\overline{d}_2} == p_{\overline{d}_a d_b\overline{d}_1\overline{d}_2} = p_{d_a\overline{d}_b\overline{d}_1\overline{d}_2} = 0$, and $p_{\overline{d}_a d_b\overline{d}_1 d_2} = p_{\overline{d}_a d_b d_1\overline{d}_2} = p_{d_a\overline{d}_b\overline{d}_1 d_2} = p_{d_a\overline{d}_b d_1\overline{d}_2} = \frac{1}{4}$.

As with the destructive measurements, we have inconsistencies between the two complementary experimental conditions. This shows that which-path information creates a context that is different from the one leading to interference. In other words, the joint probabilities obtained in the non-destructive measurement are once again incompatible with the marginals for the interference patterns contained in Case 5.

In both types of experiments, described by Cases 1 through 8, the incompatibility of contexts is reflected in the non-existence of a joint (quasi) negative probability distribution for all possible outcomes. This reflects the strong contextuality of each setup, interference or which-path, leading to observables D_1 and D_2 that are contextuality biased. It is interesting at this point to notice that this would represent, in a trivial way, a case where an experimenter could choose to observe or not D_A or D_B, and such observation would change the probabilities in D_1 and D_2. Thus, in a trivial sense, the observation of, say, D_A or not could be used to *signal* another experimenter at D_1. Though this is *not* what is usually called signaling in the literature, as it does not involve any spacelike separations between a transmitter and a receiver, it does clarify the relationship between the absence of contextual bias and the no-signalling condition. To distinguish this, Kofler and Brukner (2013) coined the term *signaling in time,* but here we use the term *contextual measurement biases* suggested by Dzhafarov and Kujala (2014d). See Oas et al. (2014); Dzhafarov and Kujala (2014d) for a somewhat more detailed discussion of this point, including the relationship between the existence of probability distributions (including negative ones) and signaling.

Case 5-8 are equivalent to the spirit of Feynman's discussions about the double-slit in his 1987 paper, and has been experimentally realized by Scully et al. (1994). Even though Scully et al. (1994) had access to the outcomes of Case 6, 7, and 8 to infer the joint probability distribution, they used counterfactual reasoning to compute a negative probability distribution that was consistent with Case 5. To do so, they had to discard certain measurements from their marginals, say, by only looking at cases where no detection happened at detector B in case 7. Let us examine the details of this counterfactual computations. First one needs to determine what are the actual observable conditions that constrain the marginal distributions. If we put detectors on both paths, we observe

$$P(d_a d_b) = 0, \tag{4.4}$$

$$P(\overline{d_a} d_b) = \frac{1}{2}, \tag{4.5}$$

$$P(d_a \overline{d_b}) = \frac{1}{2}, \tag{4.6}$$

and

$$P(\overline{d_a d_b}) = 0, \tag{4.7}$$

which corresponds to having only one photon at a time.

Whenever we observe in detector D_1 we do not observe in D_2, and *vice versa*. Furthermore, since we have a single photon, we never observe in both detectors or in neither. Finally, interference requires that we only observe in D_1. Therefore,

$$P(d_1 \overline{d}_2) = 1 \tag{4.8}$$

$$P(\overline{d}_1 d_2) = P(d_1 d_2) = P(\overline{d}_1 \overline{d}_2) = 0. \tag{4.9}$$

Now for what Feynman considered the disturbing issue. If we put a detector in arm A or B, from (4.4)-(4.7) we can "infer" that whenever we observe the particle *not* being in A, then the particle must be (probability 1) in B. But when we block the path, the probabilities are

$$P(\overline{d_a} d_1 \overline{d}_2) = P(\overline{d_a \overline{d}_1} d_2) = \frac{1}{2}, \tag{4.10}$$

and

$$P(\overline{d_b} d_1 \overline{d}_2) = P(\overline{d_b \overline{d}_1} d_2) = \frac{1}{2}. \tag{4.11}$$

The "disturbing" aspect comes from the nonmonotonicity of the above probabilities. How can $P(\overline{d}_1 d_2) = 0$, according to (4.9), while $P(\overline{d_a \overline{d}_1} d_2) = \frac{1}{2}$, from (4.10), given that $\overline{d_a \overline{d}_1} d_2$ is a proper subset of all events where $\overline{d}_1 d_2$?

Of course, this nonmonotonic property cannot be reproduced by Kolmogorov's axioms. To see this, let S_1 and S_2 be two sets in $\mathcal{F}$ such that $S_1 \subseteq S_2$ (as is the case for $S_1 = \{\omega_i \in \Omega | \mathbf{D}_1 = -1, \mathbf{D}_2 = 1\}$ and $S_2 = \{\omega_i \in \Omega | \mathbf{A} = -1, \mathbf{D}_1 = -1, \mathbf{D}_2 = 1\}$ above). Then, we can construct a set $S_1' = S_2 \backslash S_1$ such that $S_1 \cup S_1' = S_2$ and $S_1 \cap S_1' = \emptyset$. From K3 we have that $p(S_1 \cup S_1') = p(S_1) + p(S_1') = p(S_2)$, and from K1 we have at once that $p(S_1) \leq p(S_2)$ if $S_1 \subseteq S_2$, which is clearly violated by the probabilities above. Notice that in order to prove monotonicity, we had to use the non-negativity axiom K1. However, since negative probabilities violate K1, they may be nonmonotonic. For instance, from $P(S_1 \cup S_1') = P(S_1) + P(S_1') = p(S_2)$, it is possible to have $P(S_1) > P(S_2)$ if $P(S_1') < 0$.

Before we compute the joint (quasi) negative probabilities from the assumptions above, let us examine in more detail (4.10) and (4.11). The fact that each add to one corresponds to a selection of experiments where no detection happens on D_A or D_B. In other words, we are only looking at a subset of all possible experimental outcomes (essentially, this is equivalent to a postselection of data). In fact, we can see that (4.10) and (4.11) are distinct from what one observes in Case 2 and Case 3 or Case 6 and Case 7, which, as we pointed out earlier, are incompatible with Case 1 or Case 5, respectively. In this restricted data set, the counterfactual reasoning leads to a weaker context-dependency between variables, allowing for the existence of a joint negative probability distribution, as we now show.

From (4.10) and (4.11) we obtain the following set of linear equations

$$P\left(\overline{d_a}\cdot d_1\overline{d}_2\right) = P\left(\overline{d_a}d_bd_1\overline{d}_2\right) + P\left(\overline{d_ad_b}d_1\overline{d}_2\right) = \frac{1}{2}, \tag{4.12}$$

$$P\left(\overline{d_a}\cdot\overline{d}_1d_2\right) = P\left(\overline{d_a}d_b\overline{d}_1d_2\right) + P\left(\overline{d_ad_bd_1}d_2\right) = \frac{1}{2}, \tag{4.13}$$

$$P\left(\cdot\overline{d_bd_1}d_2\right) = P\left(d_a\overline{d_bd_1}d_2\right) + P\left(\overline{d_ad_bd_1}d_2\right) = \frac{1}{2}, \tag{4.14}$$

and

$$P\left(\cdot\overline{d_b}d_1\overline{d}_2\right) = P\left(d_a\overline{d_b}d_1\overline{d}_2\right) + P\left(\overline{d_ad_b}d_1\overline{d}_2\right) = \frac{1}{2}. \tag{4.15}$$

From (4.8)–(4.9), we also obtain that

$$P(\cdot\cdot d_1d_2) = P(d_ad_bd_1d_2) + P\left(\overline{d_a}d_bd_1d_2\right) + P\left(d_a\overline{d_b}d_1d_2\right) + P\left(\overline{d_ad_b}d_1d_2\right) = 0, \tag{4.16}$$

$$P\left(\cdot\cdot d_1\overline{d}_2\right) = P\left(d_ad_bd_1\overline{d}_2\right) + P\left(\overline{d_a}d_bd_1\overline{d}_2\right) + P\left(d_a\overline{d_b}d_1\overline{d}_2\right) + P\left(\overline{d_ad_b}d_1\overline{d}_2\right) = 1, \tag{4.17}$$

$$P\left(\cdot\cdot\overline{d}_1d_2\right) = P\left(d_ad_b\overline{d}_1d_2\right) + P\left(\overline{d_a}d_b\overline{d}_1d_2\right) + P\left(d_a\overline{d_bd_1}d_2\right) + P\left(\overline{d_ad_bd_1}d_2\right) = 0, \tag{4.18}$$

$$P\left(\cdot\cdot\overline{d}_1\overline{d}_2\right) = P\left(d_ad_b\overline{d}_1\overline{d}_2\right) + P\left(\overline{d_a}d_b\overline{d}_1\overline{d}_2\right) + P\left(d_a\overline{d_bd_1}\overline{d}_2\right) + P\left(\overline{d_ad_bd_1}\overline{d}_2\right) = 0. \tag{4.19}$$

Finally, (4.4)–(4.7) yields

$$P(d_ad_b\cdot\cdot) = P(d_ad_bd_1d_2) + P\left(d_ad_b\overline{d}_1d_2\right) + P\left(d_ad_bd_1\overline{d}_2\right) + P\left(d_ad_b\overline{d}_1\overline{d}_2\right) = 0, \tag{4.20}$$

$$P\left(d_a\overline{d_b}\cdot\cdot\right) = P\left(d_a\overline{d_b}d_1d_2\right) + P\left(d_a\overline{d_bd_1}d_2\right) + P\left(d_a\overline{d_b}d_1\overline{d}_2\right) + P\left(d_a\overline{d_bd_1}\overline{d}_2\right) = \frac{1}{2}, \tag{4.21}$$

(4.22)

$$P\left(\overline{d_a} d_b \cdot\cdot\right) = P\left(\overline{d_a} d_b d_1 d_2\right) + P\left(\overline{d_a} d_b \overline{d_1} d_2\right) + P\left(\overline{d_a} d_b d_1 \overline{d_2}\right) + P\left(\overline{d_a} d_b \overline{d_1}\,\overline{d_2}\right) = \frac{1}{2},$$

(4.23)

$$P\left(\overline{d_a d_b} \cdot\cdot\right) = P\left(\overline{d_a d_b} d_1 d_2\right) + P\left(\overline{d_a d_b d_1} d_2\right) + P\left(\overline{d_a d_b} d_1 \overline{d_2}\right) + P\left(\overline{d_a d_b d_1}\,\overline{d_2}\right) = 0.$$

As a last condition, from axiom N2, all probabilities of elementary events must add to one, i.e.,

$$\begin{aligned}(4.24)\; & P\left(d_a d_b d_1 d_2\right) + P\left(d_a d_b d_1 \overline{d_2}\right) + P\left(d_a d_b \overline{d_1} d_2\right) + P\left(d_a d_b \overline{d_1}\,\overline{d_2}\right) + \\ & P\left(d_a \overline{d_b} d_1 d_2\right) + P\left(a\overline{d_b} d_1 \overline{d_2}\right) + P\left(a\overline{d_b d_1} d_2\right) + P\left(a\overline{d_b d_1}\,\overline{d_2}\right) + \\ & P\left(\overline{d_a} d_b d_1 d_2\right) + P\left(\overline{d_a} d_b d_1 \overline{d_2}\right) + P\left(\overline{d_a} d_b d_1 d_2\right) + P\left(\overline{d_a} d_b \overline{d_1}\,\overline{d_2}\right) + \\ & P\left(\overline{d_a d_b} d_1 d_2\right) + P\left(\overline{d_a d_b} d_1 \overline{d_2}\right) + P\left(\overline{d_a d_b d_1} d_2\right) + P\left(\overline{d_a d_b d_1}\,\overline{d_2}\right) \quad = \quad 1\end{aligned}$$

The general solution to the underdetermined (and not independent) system of equations (4.12)–(4.24) is

(4.25)

$$\begin{array}{rr} P\left(d_a d_b d_1 d_2\right) = \alpha, & P\left(d_a d_b d_1 \overline{d_2}\right) = \theta + \frac{1}{2}\left(\delta - \gamma + \beta - \alpha\right), \\ P\left(d_a d_b \overline{d_1} d_2\right) = -\frac{1}{2} - \theta, & P\left(d_a d_b \overline{d_1}\,\overline{d_2}\right) = \frac{1}{2} + \frac{1}{2}(-\delta + \gamma - \beta - \alpha), \\ P\left(d_a \overline{d_b} d_1 d_2\right) = \frac{1}{2}(-\delta - \gamma + \beta - \alpha), & P\left(d_a \overline{d_b} d_1 \overline{d_2}\right) = \frac{1}{2} - \theta + \frac{1}{2}(-\delta + \gamma - \beta + \alpha) \\ P\left(d_a \overline{d_b d_1} d_2\right) = \theta, & P\left(d_a \overline{d_b d_1}\,\overline{d_2}\right) = \delta, \\ P\left(\overline{d_a} b d_b d_1 d_2\right) = \frac{1}{2}(\delta - \gamma - \beta - \alpha), & P\left(\overline{d_a} d_b d_1 \overline{d_2}\right) = \frac{1}{2} - \theta + \frac{1}{2}(-\delta + \gamma - \beta + \alpha), \\ P\left(\overline{d_a} d_b \overline{d_1} d_2\right) = \theta, & P\left(\overline{d_a} d_b \overline{d_1}\,\overline{d_2}\right) = \beta, \\ P\left(\overline{d_a d_b} d_1 d_2\right) = \gamma, & P\left(\overline{d_a d_b} d_1 \overline{d_2}\right) = \theta + \frac{1}{2}(\delta - \gamma + \beta - \alpha), \\ P\left(\overline{d_a d_b d_1} d_2\right) = \frac{1}{2} - \theta, & P\left(\overline{d_a d_b d_1}\,\overline{d_2}\right) = -\frac{1}{2} + \frac{1}{2}(-\delta - \gamma - \beta + \alpha), \end{array}$$

where α, β, γ, δ, and θ are arbitrary constants. It is clear from (4.25) that no nonnegative solution exists for (4.12)–(4.24). Furthermore, because the system is underdetermined, there are an infinite number of solutions that satisfy (4.12)–(4.24). To find the negative probabilities, though, we need to minimize the L1 norm, M^*. Doing so for (4.25) is straightforward but tedious, and we can show that such minimum happens when $0 \leq \alpha \leq \frac{1}{2}$, $\beta = 0$, $\delta = 0$, $\theta = 0$, and $\alpha = -\gamma$. This gives us the general solution minimizing M^* as

$$(4.26) \qquad \begin{array}{rr} P\left(d_a d_b d_1 d_2\right) = \alpha, & P\left(d_a d_b d_1 \overline{d_2}\right) = 0, \\ P\left(d_a d_b \overline{d_1} d_2\right) = -\frac{1}{2}, & P\left(d_a d_b \overline{d_1}\,\overline{d_2}\right) = \frac{1}{2} - \alpha, \\ P\left(d_a \overline{d_b} d_1 d_2\right) = 0, & P\left(d_a \overline{d_b} d_1 \overline{d_2}\right) = \frac{1}{2}, \\ P\left(d_a \overline{d_b d_1} d_2\right) = 0, & P\left(d_a \overline{d_b d_1}\,\overline{d_2}\right) = 0, \\ P\left(\overline{d_a} d_b d_1 d_2\right) = 0, & P\left(\overline{d_a} d_b d_1 \overline{d_2}\right) = \frac{1}{2}, \\ P\left(\overline{d_a} d_b \overline{d_1} d_2\right) = 0, & P\left(\overline{d_a} d_b \overline{d_1}\,\overline{d_2}\right) = 0, \\ P\left(\overline{d_a d_b} d_1 d_2\right) = -\alpha, & P\left(\overline{d_a d_b} d_1 \overline{d_2}\right) = 0, \\ P\left(\overline{d_a d_b d_1} d_2\right) = \frac{1}{2}, & P\left(\overline{d_a b d_b d_1}\,\overline{d_2}\right) = -\frac{1}{2} + \alpha, \end{array}$$

$0 \leq \alpha \leq \frac{1}{2}$, clearly showing that $M^* = 3$. We should notice that this value of M^* is greater than the $M^* = 2$ for the Bell-EPR case (Oas et al., 2014), perhaps already suggesting that the double-slit is more contextual (see de Barros et al. (2014) for a discussion of M^* as a measure of contextuality). This stronger contextuality probably comes from the use of triple moments in the Mach-Zehnder as opposed to only pairwise two-moments in the case of the standard Bell-EPR setup.

Now that we have a negative probability distribution, we can use it to compute conditional probabilities based on the previous counterfactual assumptions. For instance, a standard question is this: if a photon is detected on D_1, what is the probability that this photon went through A and B? Using

$$P(d_a|d_1) \quad = \quad \frac{1}{N}\left[P\left(d_a d_b d_1 d_2\right)+P\left(d_a \overline{d_b} d_1 d_2\right)+P\left(d_a d_b d_1 \overline{d}_2\right)+P\left(d_a \overline{d_a} d_1 \overline{d}_2\right)\right],$$

where

$$\begin{aligned} N \quad = \quad & P\left(d_a d_b d_1 d_2\right)+P\left(\overline{d_a} d_b d_1 d_2\right)+P\left(d_a \overline{d_b} d_1 d_2\right)+P\left(d_a d_b d_1 \overline{d}_2\right) \\ & +P\left(\overline{d_a d_b} d_1 d_2\right)+P\left(\overline{d_a} d_b d_1 \overline{d}_2\right)+P\left(d_a \overline{d_b} d_1 \overline{d}_2\right)+P\left(\overline{d_a d_b} d_1 \overline{d}_2\right), \end{aligned}$$

and (4.26) we have that

$$P(d_a|d_1) \quad = \quad \frac{1}{2}+\alpha,$$

and

$$\frac{1}{2} \leq P\left(d_a|d_1\right) \leq 1.$$

Similarly,

$$P(d_b|d_1) \quad = \quad \frac{1}{N}\left[P\left(d_a d_b d_1 d_2\right)+P\left(\overline{d_a} d_b d_1 d_2\right)+P\left(d_a d_b d_1 \overline{d}_2\right)+P\left(\overline{d_a} d_b d_1 \overline{d}_2\right)\right],$$

where

$$\begin{aligned} N \quad = \quad & P\left(d_a d_b d d_2\right)+P\left(d_a \overline{d_b} d_1 d_2\right)+P\left(\overline{d_a} d_b d_1 d_2\right)+P\left(d_a d_b d_1 \overline{d}_2\right) \\ & +P\left(\overline{d_a d_b} d_1 d_2\right)+P\left(d_a \overline{d_b} d_1 \overline{d}_2\right)+P\left(\overline{d_a} d_b d \overline{d}_2\right)+P\left(\overline{d_a d_b} d_1 \overline{d}_2\right), \end{aligned}$$

and we get

$$P(d_b|d_1) \quad = \quad \frac{1}{2}+\alpha,$$

the same value we got for the conditional $P(d_b|d_1)$. For d_2, the conditional probability is not defined, as the probability for d_2 from the joint is zero. However, if we set the interferometer such that the probability of d_2 is not zero, but close to it, then $P(b|d_2)$ can be shown to approach $P(d_b|d_2) = -\frac{1}{2}+\alpha$. If that is the case, it is reasonable to assume that $\alpha = 1/2$, such that we do not have negative probabilities for d_b (conditioned on d_2). If we do so, we reach the interesting conclusion that both d_b and d_a have probability 1 given an observation on d_1. In other words, if we use the counterfactual reasoning from negative probabilities, we reach the conclusion, as Feynman often said, that the particle goes through *both* paths simultaneously.

We now end this section with a discussion of some well-known uses of counterfactual reasoning in quantum mechanics and their relationship to our discussion above. First, it is worth mentioning that the famous Leggett and Garg (1985) setup can be thought of as similar to our double-slit experiment (Kofler and Brukner, 2013). To see this, we recall that in Leggett and Garg (LG), measurements in three distinct times can coded by three ± 1-valued random variables, say $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ (Bacciagaluppi, 2014; Dzhafarov and Kujala, 2014d; de Barros et al., 2014; Dzhafarov and Kujala, 2014a). In an analogy with the double-slit, and following Kofler and Brukner (2013), we can think of $\mathbf{X}$ as a measurement of position before BS_1, $\mathbf{Y}$ as a measurement of which path (say, with $\mathbf{Y} = 1$ corresponding to A and $\mathbf{Y} = -1$ to B), and $\mathbf{Z}$ corresponding to a detection in either D_1 (for $\mathbf{Z} = 1$) or D_2 ($\mathbf{Z} = -1$). This case would correspond to contextual bias, and would not include counterfactual reasoning. However, in the original LG paper, counterfactual reasoning happens by not measuring $\mathbf{Y}$, but instead making inferences about $\mathbf{Y}$ from an absence of detection in one of the paths. As we mentioned above, such contextual bias is not surprising, as the effect of measuring could be thought as interfering with the experimental conditions themselves. This is similar to what happens in our analysis above.

Something analogous happens with the argument given by Scully et al. (1994). In his paper, he talks about negative probabilities, and shows that they lead to the interference between two possible modes. However, it is easy to see that the negative probabilities so obtained are only existent because of the same type of counterfactual reasoning shown above. That should be clear by the fact that, in their experiment, the interference pattern is existent, and therefore we have observational contextual bias, similar to the Leggett-Garg setup.

5. Final Remarks

In this paper we presented a proposed theory of negative probabilities that could be used to describe non-monotonic reasoning. Such theory was shown to be equivalent, in the case when a proper probability distribution exists, to the standard Kolmogorov probability, as the requirement of minimizing the total probability mass leads to a Kolmogorovian distribution. Furthermore, in cases where no proper joint probability exists, the minimization of the total negative mass is simply a requirement that our quasi-probability distribution is as close to a proper one as possible. Such minimization is similar to the requirement with upper probabilities of minimizing the total sum of probabilities, which may exceed one.

We did not attempt to interpret negative probabilities, but instead took the approach that they constitute a bookkeeping tool that meets a minimum rationality criteria of minimization of the L1 probability norm. This criteria is perhaps not without practical consequences. For instance, in de Barros (2014), we showed that in certain cases where the pairwise correlations lead to contradictions, this minimization results in constraints to the triple moments. In addition, in Oas et al. (2014), it was shown that the minimized L1 norm is equivalent to the CHSH parameter, S, as used in measures of non-locality (Cirel'son, 1980), specifically $M^* = S/2$. Finally, the L1 norm can be thought of as a measure of contextuality for random variables, and is closely related to other measures of contextuality, at least for three and four random-variables, and for more variables it suggests possible different classifications for contextuality (de Barros et al., 2014).

Though the double-slit experiment is the archetypical in discussions of how quantum mechanics leads to violation of the laws of probability or logic (Dalla Chiara and Giuntini, 2014), it is not the simplest and most accessible example that contains the key conceptual elements relevant to the subject. In fact, the Mach-Zehnder interferometer, as we showed above, presents the same characteristics as the double-slit experiment that are relevant to conceptual discussions of quantum mechanics, without the complications associated with the details of continuous interference patterns present in the double-slit. For that reason, in our discussion of the double-slit experiment in terms of negative probabilities, we resorted to the simpler case of the Mach-Zehnder interferometer. As we saw, the Mach-Zehnder interferometer allows us to talk about the features of double-slit experiment in terms of discrete random variables, which tremendously reduce the mathematical complexity without loosing any conceptual generality.

As we showed, the two possible setups for the Mach-Zehnder interferometer, one with which-path information and another with interference, present contextual biases. This has, as a consequence, the non-existence of a joint negative (quasi) probability distribution consistent with all observations of the two Mach-Zehnder interferometer setups. This clearly corresponds to two different experimental contexts, and not only does a joint probability consistent with both contexts not exist, but no negative joint distribution exists either. This type of system, where contextuality comes from contextual biases, exhibit what one could think of as stronger contextuality than other systems, such as EPR.

This stronger contextuality may be what is reflected in the large values of M^* for the two slits, but such a connection has not been studied in detail. In fact, it is interesting to notice that the large values of M^* is associated to a set of observables that do not provide a complete picture of the experimental conditions, as it relies on counterfactuals. Perhaps there is a connection between large M^* for a restricted set of observables and contextual biases for an extended set, which could provide an interesting criteria for contextual biases. Notice that, as mentioned in Section 4, contextual biases are equivalent to the violation of the no-signaling condition in multipartite systems.

Finally, we would like to comment on Feynman's (1987) discussion of the double-slit experiment. In this paper, he argues that the non-monotonic character of quantum probabilities could be represented by non-observable negative probabilities. He then goes on and constructs (in a very informal way) a possible negative probability that could explain the outcomes of the experiment. However, as we pointed out above, negative probabilities consistent with the outcomes of the double-slit experiment are impossible, unless we make use of certain specific counterfactual reasoning. It is interesting to note that in his actual experimental realization of Feynman's double-slit experiment, Scully et al. (1994) construct a negative probability distribution, and their probabilities rely exactly on the type of counterfactual reasoning used above. Were they to try and construct a negative probability from the full range of experimental data, they would not be able to do so. A similar case is present in the LG experiment also discussed above.

Acknowledgments. This paper stems from a 2011 seminar led by the three authors, where a full review and exploration of negative probabilities was undertaken, and we profited from exchanges with the seminar participants, in particular Tom Ryckman, Claudio Carvalhaes, Michael Heaney, Michael Friedman, Hyungrok Kim, and Niklas Damiris. It was Pat Suppes's goal to establish a rigorous, consistent theory of negative probability, and this work is in line with that unfinished goal. We also benefited tremendously from collaborations and discussions with Ehtibar Dzhafarov and Janne Kujala. We also thank Guido Bacciagaluppi for discussions and for pointing us to the Kofler and Brukner reference.

References

Abramsky, S. and Brandenburger, A. (2011). The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13(11):113036.

Abramsky, S. and Brandenburger, A. (2014). An operational interpretation of negative probabilities and no-signalling models. In van Breugel, F., Kashefi, E., Palamidessi, C., and Rutten, J., editors, *Horizons of the Mind. A Tribute to Prakash Panangaden*, number 8464 in Lecture Notes in Computer Science, pages 59–75. Springer Int. Pub.

Al-Safi, S. and Short, A. (2013). Simulating all nonsignaling correlations via classical or quantum theory with negative probabilities. *Physical Review Letters*, 111(17):170403.

Aspect, A., Dalibard, J., and Roger, G. (1982). Experimental test of bell's inequalities using time- varying analyzers. *Physical Review Letters*, 49(25):1804–1807.

Aspect, A., Grangier, P., and Roger, G. (1981). Experimental tests of realistic local theories via bell's theorem. *Physical Review Letters*, 47(7):460–463.

Bacciagaluppi, G. (2014). Leggett-garg inequalities, pilot waves and contextuality. *International Journal of Quantum Foundations*, 1(1):1–17. arXiv: 1409.4104.

Bell, J. (1964). On the einstein-podolsky-rosen paradox. *Physics*, 1(3):195–200.

Bell, J. (1966). On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, 38(3):447–452.

Burgin, M. (2010). Interpretations of negative probabilities. *arXiv:1008.1287 [physics, physics:quant-ph]*. arXiv: 1008.1287.

Burgin, M. and Meissner, G. (2012). Negative probabilities in financial modeling. *Wilmott*, 2012(58):60–65.

Cirel'son, B. (1980). Quantum generalizations of bell's inequality. *Letters in Mathematical Physics*, 4(2):93–100.

Dalla Chiara, M. and Giuntini, R. (2014). The two-slits experiment, quantum logic and quantum computational logic. In Fano, V., editor, *Gino Tarozzi Philosopher of Physics. Studies in the philosophy of entanglement on his 60th birthday*. Franco Angeli, Milan, Italy.

de Barros, J. A. (2014). Decision making for inconsistent expert judgments using negative probabilities. Lecture Notes in Computer Science, pages 257–269. Springer, Berlin/Heidelberg.

de Barros, J. A., Dzhafarov, E., Kujala, J., and Oas, G. (2014). Unifying two methods of measuring quantum contextuality. *arXiv:1406.3088 [quant-ph]*. arXiv: 1406.3088.

de Barros, J. A. and Oas, G. (2014). Negative probabilities and counter-factual reasoning in quantum cognition. *arXiv:1404.3921 [physics, physics:quant-ph]*.

de Barros, J. A. and Suppes, P. (2000). Inequalities for dealing with detector inefficiencies in greenberger-horne-zeilinger type experiments. *Physical Review Letters*, 84(5):793–797.

de Barros, J. A. and Suppes, P. (2001). Probabilistic results for six detectors in a three-particle GHZ experiment. In Bricmont, J., Durr, D., Galavotti, M. C., Ghirardi, G., Petruccione, F., and Zanghi, N., editors, *Chance in Physics*, volume 574 of *Lectures Notes in Physics*, page 213.

Dzhafarov, E. and Kujala, J. (2012). Selectivity in probabilistic causality: Where psychology runs into quantum physics. *Journal of Mathematical Psychology*, 56(1):54–63.

Dzhafarov, E. and Kujala, J. (2013a). All-possible-couplings approach to measuring probabilistic context. *PLoS ONE*, 8(5):e61712.

Dzhafarov, E. and Kujala, J. (2013b). Random variables recorded under mutually exclusive conditions: Contextuality-by-default. *arXiv:1309.0962 [quant-ph]*.

Dzhafarov, E. and Kujala, J. (2014a). Generalizing bell-type and leggett-garg-type inequalities to systems with signaling. *arXiv:1407.2886 [quant-ph]*.

Dzhafarov, E. and Kujala, J. (2014b). Probabilistic contextuality in {EPR}/{B}ohm-type systems with signaling allowed. *arXiv:1406.0243 [quant-ph, q-bio]*.

Dzhafarov, E. and Kujala, J. (2014c). A qualified kolmogorovian account of probabilistic contextuality. *Lecture Notes in Computer Science*, 8369:201–212.

Dzhafarov, E. N. and Kujala, J. V. (2014d). Contextuality in generalized klyachko-type, bell-type, and leggett-garg-type systems. *arXiv:1411.2244 [physics, physics:quant-ph]*. arXiv: 1411.2244.

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47(10):777–780.

Feynman, R. (1987). Negative probability. In Hiley, B. and Peat, F., editors, *Quantum implications: essays in honour of David Bohm*, pages 235–248. Routledge, London and New York.

Feynman, R., Leighton, R., and Sands, M. (2011). *The Feynman lectures on physics: Mainly mechanics, radiation, and heat*, volume 3. Basic Books, New York, NY.

Fine, A. (1982). Hidden variables, joint probability, and the bell inequalities. *Physical Review Letters*, 48(5):291–295.

Fine, T. L. (1994). Upper and lower probability. In Humphreys, P., editor, *Patrick Suppes: Scientific Philosopher*, number 234 in Synthese Library, pages 109–133. Springer Netherlands.

Garola, C., Persano, M., Pykacz, J., and Sozzo, S. (2014). Finite local models for the GHZ experiment. *International Journal of Theoretical Physics*, 53(2):622–644.

Garola, C. and Sozzo, S. (2009). The ESR model: A proposal for a noncontextual and local hilbert space extension of QM. *EPL (Europhysics Letters)*, 86(2):20009.

Garola, C. and Sozzo, S. (2011a). Generalized observables, bell's inequalities and mixtures in the ESR model for QM. *Foundations of Physics*, 41(3):424–449.

Garola, C. and Sozzo, S. (2011b). The modified bell inequality and its physical implications in the ESR model. *International Journal of Theoretical Physics*, 50(12):3787–3799.

Greenberger, D. M., Horne, M. A., and Zeilinger, A. (1989). Going beyond {B}ell's theorem. In Kafatos, M., editor, *Bell's theorem, Quantum Theory, and Conceptions of the Universe*, volume 37 of *Fundamental Theories of Physics*, pages 69–72. Kluwer, Dordrecht, Holland.

Halliwell, J. J. and Yearsley, J. M. (2013). Negative probabilities, fine's theorem, and linear positivity. *Physical Review A*, 87(2):022114.

Hartmann, S. and Suppes, P. (2010). Entanglement, upper probabilities and decoherence in quantum mechanics. In Suárez, M., Dorato, M., and Rédei, M., editors, *EPSA Philosophical Issues in the Sciences*, pages 93–103. Springer Netherlands.

Howard, M., Wallman, J., Veitch, V., and Emerson, J. (2014). Contextuality supplies the /'magic/' for quantum computation. *Nature*, 510(7505):351–355.

Jacques, V., Wu, E., Grosshans, F., Treussart, F., Grangier, P., Aspect, A., and Roch, J.-F. (2007). Experimental realization of wheeler's delayed-choice gedanken experiment. *Science*, 315(5814):966–968.

Khrennikov, A. (1993a). p-adic probability theory and its applications. {T}he principle of statistical stabilization of frequencies. *Theoretical and Mathematical Physics*, 97(3):1340–1348.

Khrennikov, A. (1993b). p-adic statistical models. *Doklady Akademii Nauk*, 330(3):300–304.

Khrennikov, A. (1994a). Discrete {Q}p-valued probabilities. *Russian Academy of Sciences-Doklady Mathematics-AMS Translation*, 48(3):485–490.

Khrennikov, A. (1994b). *p-Adic Valued Distributions in Mathematical Physics*, volume 309 of *Mathematics and Its Applications*. Springer Science+Business Media, Dordrecht, Holland.

Khrennikov, A. (2009). *Interpretations of probability*. Walter de Gruyter, Berlin.

Kochen, S. and Specker, E. P. (1967). The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17:59–87.

Kochen, S. and Specker, E. P. (1975). The problem of hidden variables in quantum mechanics. In Hooker, C., editor, *The Logico-Algebraic Approach to Quantum Mechanics*, pages 293–328. D. Reidel Publishing Co., Dordrecht, Holland.

Kofler, J. and Brukner, Č. (2013). Condition for macroscopic realism beyond the leggett-garg inequalities. *Physical Review A*, 87(5):052115.

Kolmogorov, A. (1950). *Foundations of the theory of probability*, volume viii. Chelsea Publishing Co., Oxford, England.

Leggett, A. and Garg, A. (1985). Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Physical Review Letters*, 54(9):857–860.

Markiewicz, M., Kurzynski, P., Thompson, J., Lee, S.-Y., Soeda, A., Paterek, T., and Kaszlikowski, D. (2013). Unified approach to contextuality, non-locality, and temporal correlations. *arXiv:1302.3502 [quant-ph]*.

Oas, G., de Barros, J. A., and Carvalhaes, C. (2014). Exploring non-signalling polytopes with negative probability. *Physica Scripta*. In Press. arXiv:1404.3831 [quant-ph].

Rao, K. P. S. B. and Rao, M. B. (1983). *Theory of Charges: A Study of Finitely Additive Measures*. Academic Press.

Scully, M. O. and Druhl, K. (1982). Quantum eraser: A proposed photon correlation experiment concerning observation and "delayed choice" in quantum mechanics. *Physical Review A*, 25(4):2208–2213.

Scully, M. O., Walther, H., and Schleich, W. (1994). Feynman's approach to negative probability in quantum mechanics. *Physical Review A*, 49(3):1562–1566.

Silverman, M. P. (1995). *More than one mystery: explorations in quantum interference*. Springer-Verlag, New York.

Sozzo, S. and Garola, C. (2010). A hilbert space representation of generalized observables and measurement processes in the ESR model. *International Journal of Theoretical Physics*, 49(12):3262–3270.

Suppes, P. and de Barros, J. A. (1994a). Diffraction with well-defined photon trajectories: A foundational analysis. *Foundations of Physics Letters*, 7(6):501–514.

Suppes, P. and de Barros, J. A. (1994b). A random-walk approach to interference. *International Journal of Theoretical Physics*, 33(1):179–189.

Suppes, P., de Barros, J. A., and Oas, G. (1996a). A collection of probabilistic hidden-variable theorems and counterexamples. In Pratesi, R. and Ronchi, L., editors, *Waves, Information, and Foundations of Physics: a tribute to Giuliano Toraldo di Francia on his 80th birthday*, Florence, Italy. Italian Physical Society.

Suppes, P., de Barros, J. A., and Sant'Anna, A. S. (1996b). A proposed experiment showing that classical fields can violate bell's inequalities. *arXiv:quant-ph/9606019*.

Suppes, P., de Barros, J. A., and Sant'Anna, A. S. (1996c). Violation of bell's inequalities with a local theory of photons. *Foundations of Physics Letters*, 9(6):551–560.

Suppes, P., Sant'Anna, A., and de Barros, J. A. (1996d). A particle theory of the casimir effect. *Foundations of Physics Letters*, 9(3):213–223.

Suppes, P. and Zanotti, M. (1981). When are probabilistic explanations possible? *Synthese*, 48(2):191–199.

Suppes, P. and Zanotti, M. (1991). Existence of hidden variables having only upper probabilities. *Foundations of Physics*, 21(12):1479–1499.

Wheeler, J. A. (1978). The "past" and the "delayed-choice" double-slit experiment. *Mathematical Foundations of Quantum Theory*, page 9.

Zhu, X., Wei, Q., Liu, Q., and Wu, S. (2013). Negative probabilities and information gain in weak measurements. *Physics Letters A*, 377(38):2505–2509.

* Liberal Studies Program, San Francisco State University, San Francisco, CA, 94132., †Stanford Pre-Collegiate Studies, Stanford University,Stanford, CA, 94305-4101., ⋆ CSLI, Ventura Hall, Stanford University, Stanford, CA, 94305-4115.